1987-03

Reduction in bandwidth and buffer size by using modified Huffman coding after dropping the less frequent source symbols

Corapcioglu, Ahmet

http://hdl.handle.net/10945/22451

NAVAL POSTGRADUATE SCHOOL

Monterey, California

THESIS

REDUCTION IN BANDWIDTH AND BUFFER SIZE
BY USING THE MODIFIED
HUFFMAN CODING AFTER DROPPING THE LESS
FREQUENT SOURCE SYMBOLS

by

Ahmet Corapcioglu

March 1987

Thesis Advisor Richard W. Hamming

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY 2b DECLASSIFICATION / DOWNGRADING SCHEDULE	3 DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; Distrubution is unlimited
4 PERFORMING ORGANIZATION REPORT NUMBER(S)	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Naval postgraduate School 6b OFFICE SYMBOL (If applicable) 52	7a NAME OF MONITORING ORGANIZATION Naval Postgraduate School
6c ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000	7b ADDRESS (City, State, and ZIP Code) Monterey, California 93943-5000
8a NAME OF FUNDING / SPONSORING ORGANIZATION 8b OFFICE SYMBOL (If applicable)	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code)	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
Reduction in bandwidth & buffer size by using Modified Huffman Coding after dropping the less frequent source symbols.

12 PERSONAL AUTHOR(S)
Corapcioglu, Ahmet

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Master's Thesis	FROM ______ TO ______	1987 March	123

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Bandwidth, Buffer, Huffman Coding, Modification, Average Code Length, Variance.

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

This research employs the modified Huffman coding after dropping the selected source symbols, with respect to their usage frequencies. The expected results are a decrease in average length as well as in variance. The ultimate purpose is to reduce the required bandwidth and the buffer size in a communications system.

20 DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Richard W. Hamming	22b TELEPHONE (Include Area Code) 408-646-2655 22c OFFICE SYMBOL 52Hg

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Reduction in Bandwidth and Buffer Size by Using The Modified Huffman Coding After Dropping The Less Frequent Source Symbols

by

Ahmet Corapcioglu
Lieutenant J.G., Turkish Navy
B.S., Turkish Naval Academy, 1979

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN TELECOMMUNICATIONS SYSTEMS MANAGEMENT

from the

NAVAL POSTGRADUATE SCHOOL
March 1987

ABSTRACT

This research employs the modified Huffman coding technique after dropping the selected source symbols, with respect to their usage frequencies. The expected results are a decrease in average length as well as in variance. The ultimate purpose is to reduce the required bandwidth and buffer size in a communications system.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

A. INTRODUCTION

In communications systems, the limited available bandwidth forces communications engineers and managers to use the existing available bandwidth effectively. The advantages of digital communications techniques, explained in Chapter 2, convince the communications world to switch from analog to digital communications. However, digital techniques require more bandwidth than analog techniques. In addition to bandwidth requirements, the lack of necessary bandwidth requires the untransmitted digits to be stored in buffers until their turn for transmission occurs. Hence, the less bandwidth that is available for the communications system, the larger the buffer size required.

One solution to this problem is to employ a variable length coding technique, known as Huffman coding. In Huffman coding, the code words are assigned to each source alphabet symbol with respect to its usage frequency in the language. The more frequent symbols are assigned shorter code words, and vice versa.

In [Ref. 1,] the Huffman coding process was modified by employing two modification parameters, N, and E. Modification of Huffman coding results in a smaller increase in average code length, with a larger decrease in variance. In [Ref. 2,] an additional modification parameter, K, was introduced. In his research, Akinsel concluded that the parameter E was the most robust. Both authors, after modification, calculated the reduction in bandwidth and buffer size by comparing Modified Huffman coding results with Block coding results.

In this research, source symbols are also encoded by using the Modified Huffman coding technique. Modification is done only by employing parameter E, since it is the most effective of the three. The main difference of this research is to drop the less frequent source symbols before encoding the messages. The anticipated results are a reduction in average length in addition to a reduction in variance. Further, a reduction in the required transmission bandwidth, as well as the buffer size, is expected. The same idea is examined by dropping the more frequent source symbols and dropping a combination of more or less frequent symbols.

B. STRUCTURE

The structure of the remainder of the thesis is as follows.

Chapter 2 discusses advantages and disadvantages of digital communications, presents a brief background in Huffman coding, modification, and introduces the idea of dropping symbols.

In Chapter 3, the Turkish alphabet is encoded before and after dropping the less frequent source symbols. The effect of the dropping process and the modification parameter E on average length and variance are observed.

The effect of dropping the less frequent symbols on the meaning of the messages is examined in Chapter 4.

Chapter 5 compares bandwidth and buffer size requirements with Block coding, Huffman coding, Modified Huffman coding, and the dropping process, by using the simulation model of the communication systems.

In Chapter 6, two other alternatives, dropping the more frequent and the more and less frequent symbol combinations, are briefly explained.

A summary of results and conclusions are provided in Chapter 7.

II. BACKGROUND

A. WHY DIGITAL?

In his book, K. Feher [Ref. 3] states that, "at the present time, most major operational terresterial line - of - sight and satellite microwave systems use analog FM modulation techniques. However, the trend in new development is such that the overwhelming majority of new microwave systems will employ digital methods."

This comment is not true just of microwave communications systems, but also true of all types of communications systems. In the communications engineering field, most of the new developments employ digital techniques, such as digital signal processing, digital multiplexing, digital switching, transmission techniques, etc. [Ref. 3.] Hence, the solution to the communications requirements is mostly satisfied with these digital approaches. The new additions to the existing communications networks tend to use digital techniques. This trend, with the support of the new developments in the digital area, will increase and by the end of this century almost all of the new solutions for the communications requirements will be digital.

A major advantage of digital communications is low signal-to-noise ratio [Ref. 4.] In the analog communications case, all kinds of undesirable amplitude, frequency and phase variations, caused by either external source noise or systems hardware, are received at the receiver.

However, in digital transmission, while the digital pulses are also affected by the same sources, the receiver extracts the original information simply by looking at "whether the received signal at the receiver at the time of sampling is either above or below a particular voltage threshold" [Ref. 4.]

The required distance between stations should not be more than six thousand feet. For longer distance communications needs, repeaters are used very effectively. Repeaters can, after detecting the bit - pattern of the signal, reconstruct and transmit the signal without any error, either to the destination receiver or to another repeater.

In addition to this major advantage, digital communications networks have more advantages and, like all the other real life systems have some disadvantages. These advantages and disadvantages can be summarized as follows [Ref. 5.]

1. Advantages Of The Digital Network

a. Ease of multiplexing

In digital communications, mostly Time Division Multiplexing (TDM) is used. Although time division multiplexing of analog signals is possible, "the vulnerability of narrow analog pulses to noise, distortion, crosstalk and intersymbol interference " [Ref. 5] makes this option useless. So, for the multiplexing of the analog signals, Frequency Division Multiplexing (FDM) is commonly used. The TDM equipment cost is less expensive than the FDM equipment cost.

b. Ease of Signaling

In digital systems, control information (on hook/off hook, address digits, etc.) can be inserted and extracted from a message stream independently of the transmission medium. So, the transmission system can be designed separately from the transmission medium. Taking this one step further, control functions and their formats can be modified independently of the transmission subsystem. The system upgrading can be done without any impact on the control modules at either end of the link.

The analog transmission systems also require special attention for control signaling. Many of the different analog systems require unique control signals. The control formats depend on the nature of both the transmission system and its terminal equipment. Additionally, in the interfaces between different subsystems, this unique control format requires the conversion of the control signals from one format to another.

c. Use of Modern Technology

Logic gates and memory, as they are used in digital computer technology, can easily be used in digital signaling. The main idea in digital switching is simply to use the "AND gate with one logic input assigned to the message signal and the other inputs used for control" [Ref. 5: p.66]. Hence, the same technological development in the computer logic circuits can be applied in digital integrated circuit technology.

Large scale integrated circuits (LSI) are developed specifically for telecommunications. LSI chips improve the cost - effectiveness, the size and the reliability of the communications system that uses digital techniques. Despite the currently common usage of frequency division multiplexing access (FDMA), techniques indicate that future satellite communications will be digital.

In fiber optic communications, the interface of electronics and optical fibers uses primarily the "on-off" mode of operations. Hence, the transmission link itself emphasizes the digital mode of operation.

In the digital signal transmission area, which can be used for transmitting both analog and digital wave forms, the digital technique is used.

d. Integration of Transmission and Switching

In analog phone systems, the transmission and multiplexing equipment are considered separately, because they are functionally independent. However, in digital systems, since the TDM of the signal is very similar to the time division switching function, both can easily be integrated. The benefits of this integration are:

(1) demultiplexing equipment at switching offices is unnecessary

(2) greatly improved end - to - end voice quality

(3) cable entrance requirements and mainframe distribution of wire pairs are reduced.

e. Signal Regeneration

The analog wave form is transformed into a sequence of discrete values. These discrete values are then represented by a sequence of binary digits. In the transmission each binary digit is represented by one of the two possible signal values, such as "pulse and no pulse" or "positive pulse and negative pulse," etc. At the receiver, regeneration of the original signal is very simple, since all that is needed is to distinguish one of these two possible signal values. If the transmission distance is not very long, the effect of the external noise source on the transmitted signal will not be enough to change its value outside the threshold values. If the distance is longer than the required distance the undesired noise will be strong enough to destroy the signal. In order to overcome this difficulty, repeaters are stationed between the source and the destination. They detect the bit pattern and reconstruct and transmit the pattern to the next repeater.

f. Ease of Encryption

Decoding and encoding of a digital bit stream is much easier than that of an analog signal. This characteristic, especially for military applications of digital techniques, makes it very attractive.

2. Disadvantages of Digital Network

a. Analog - to - Digital Conversion

One of the main expenses of the digital network is the conversion cost. Since most of the digital networks use existing analog networks, the savings due to reduced equipment, such as multipliers and switches, generally covers the conversion cost.

b. *Need for Time Synchronization*

Although some transmission systems in analog networks require some sort of synchronization, like FDM transmission system carrier synchronization, synchronization in analog networks is not a requirement. It can be considered as a function of the transmission system. On the other hand, in digital networks time synchronization is a requirement for optimum detection at the receiver, "the sample clock must be synchronized to the pulse arrival time" [Ref. 5.] This problem increases as the number of digital transmission links and switches in the network increases.

c. *Increased Bandwidth*

In a digital system a waveform is sampled and these samples are coded into binary digits. For each digit, one individual pulse is transmitted. In the analog systems, the transmission of a waveform does not require more bandwidth than the underlying original wave. Hence, it can easily be seen that digital systems require more bandwidth than analog systems.

Transmission quality is reached by representing the waveform with more digits. This increases the bandwidth. The bandwidth increase, in voice digitization, is directly dependent on the form of coding or modulation used.

When we look at the existing local analog loop, since the bandwidth is underutilized, an increase in bandwidth due to digitization might not create a big problem. In long - haul systems, since the bandwidth utilization is high, an increase in bandwidth is less acceptable.

One of the ways to overcome this major disadvantage of additional bandwidth requirements in digital communications is to reduce the bandwidth requirement by employing variable length encoding techniques.

Contrary to the block coding technique, which is a fixed bit sequence length, Huffman coding assigns the bit sequence to each source symbol according to its frequency of occurence in the source alphabet. The higher frequency symbol is assigned shorter bit sequence and vice versa [Ref. 6.]

Two important requirements of this coding technique can be stated as follows:

(1) Each character should be coded with a unique bit sequence

(2) Decoding should be done in a way that the beginning and end of each character is known without any special indicator [Ref. 7.]

As explained in the next section, although Huffman coding reduces the average length of code words, it introduces the variance of the code length. Reduction in average code length results in a gain (reduction) in bandwidth. On the other hand, a high variation on the code length requires a larger buffer in the system.

One possible solution presented in this thesis is encoding messages by using Modified Huffman coding (with modification parameters) after dropping the less frequent source symbols. The expected result is smaller average length and also smaller variance than either block coding and Huffman coding would provide.

B. HUFFMAN CODING

Huffman coding is a minimum - redundancy code which uses the frequencies of the symbols for assigning the binary digits in encoding. The more frequent (probable) symbol will have the shorter length encoding [Ref. 7.] Let's assume that there are N symbols in the message. P_i is the probability of the ith symbol where, i = 1,...,N. So,

$$\sum P_i = 1.$$

L_i is the length of ith encoding. The average length of the code is [Ref. 6:]

$$L_{ave} = \sum P_i L_i$$

We can rewrite the symbols according to their probabilities in decreasing order:

$$P_1 \geq P_2 \geq P_3 \geq ... \geq P_N$$

and, for an optimum code (minimum - redundancy code), lengths in increasing order:

$$L_1 \leq L_2 \leq L_3 \leq ... \leq L_N$$

The Huffman binary coding procedure begins with arranging the symbols in order of decreasing probabilities. Then, the two least probable symbols are combined into one symbol. The new symbol's probability is the sum of the two least probable symbols' probabilities. The new symbol is placed in decreasing probability order. And again the two least probable symbols are combined. The process is repeated until we have just two symbols remaining. At that point we can assign the codes to the

symbols. For the sake of an example, let's assign 0 to the upper symbol and 1 to the lower symbol. Consider the following example [Ref. 2.]

We have our source alphabet and symbol probabilities. First we write them in decreasing order. See Figure 2.1.

Symbol	Probability
S1	0.4
S2	0.2
S3	0.2
S4	0.1
S5	0.05
S6	0.05

Figure 2.1 Source Alphabet And Symbol Probabilities.

Now we combine the two least probable symbols (S5, S6) and place the new symbol into decreasing order (Figure 2.2.) We keep combining the two least probable symbols until we have just two symbols (Figure 2.3.) The least probable symbols are shown with #, and the combined one with *.

Symbol	p	p
S1	0.4	0.4
S2	0.2	0.2
S3	0.2	0.2
S4	0.1	0.1
S5	0.05#	*0.1
S6	0.05#	

Figure 2.2 Reduction Process.

We assign (0) and (1) to the last two symbols. In our example, 1 is assigned to the upper symbol and 0 is assigned to the lower symbol. Now we are ready to trace backwards until we assign codes to each symbol in the alphabet. On the way back, the combined symbols expand into two branchs. We add one more digit for each branch. Figure 2.4 shows the splitting process and Figure 2.5 shows the assigned code words.

Symbol	P	P	P	P	P
S1	0.4	0.4	0.4	0.4	*0.6
S2	0.2	0.2	0.2	0.4#	0.4
S3	0.2	0.2	0.2#	*0.2#	
S4	0.1	0.1#	*0.2#		
S5	0.05#	*0.1#			
S6	0.05#				

Figure 2.3 Huffman Coding Reduction Process.

Symbol	P	P	P	P	P
S1	0.4(1)	0.4(1)	0.4(1)	0.4(1)	*0.6(0)
S2	0.2(01)	0.2(01)	0.2(01)	*0.4(00)#	0.4(1)
S3	0.2(000)	0.2(000)	0.2(000)#	0.2(01)#	
S4	0.1(0010)	0.1(0010)#	*0.2(001)#		
S5	0.05(00110)#	*0.1(0011)#			
S6	0.05(00111)#				

Figure 2.4 Splitting Process.

C. MODIFICATION OF HUFFMAN CODING

In the example given in section B, the combined symbols are placed as low as possible in the decreasing probabilities order. Using the given symbols probabilities and final code length of each symbol (Figure 2.5), we can calculate the average code length as follows.

Average code length :

$$L = \sum P_i L_i \qquad \text{so}$$

$$L = (0.4)1 + (0.2)2 + (0.2)3 + (0.1)4 + (0.05)5 + (0.05)5$$

$$L = 2.3$$

Symbol	Code Word
S1	1
S2	01
S3	000
S4	0010
S5	00110
S6	00111

Figure 2.5 Final Code Words.

and the variance is

$$V = 0.4(1 - 2.3)^2 + 0.2(2 - 2.3)^2 + 0.2(3 - 2.3)^2 +$$
$$0.1(4 - 2.3)^2 + 0.05(5 - 2.3)^2 + 0.05(5 - 2.3)^2$$
$$V = 1.81$$

On the other hand, if we place the combined symbol as high as possible in the decreasing probability order, see Figure 2.6, we will have different lengths for source symbols (2, 2, 2, 3, 4, 4). The average length and the variance are :

$$L = 0.4(2) + 0.2(2) + 0.2(2) + 0.1(3) + 0.05(4) + 0.05(4)$$
$$L = 2.3$$
$$V = 0.4(2 - 2.3)^2 + 0.2(2 - 2.3)^2 + 0.2(2 - 2.3)^2$$
$$+ 0.1(3 - 2.3)^2 + 0.05(4 - 2.3)^2 + 0.05(4 - 2.3)^2$$
$$V = 0.41$$

Although the average lengths are the same, placing the combined symbols as high as possible gives us a smaller variance. This is a desirable result for communications systems. We want to have smaller length codes with small variances than block coding provides.

As explained in [Ref. 1] and [Ref. 2,] by employing three parameters K, N, and E, we can achieve lower variance Huffman codes, with a higher average length. The initial decrease in variance is much more than the increase in the average length, so the system will have a marginal profit.

1. **Parameters**

a. Parameter N

N is an integer value. Instead of placing the combined symbol in decreasing order with respect to its probability, is is placed in a higher position according to the

Symbol	P	P	P	P	P
S1	0.4(00)	0.4(00)	0.4(00)	*0.4(1)	*0.6(0)
S2	0.2(10)	0.2(10)	0.2(01)	0.4(00)#	0.4(1)
S3	0.2(11)	0.2(11)	*0.2(10)#	0.2(01)#	
S4	0.1(010)	*0.1(010)#	0.2(11)#		
S5	0.05(0110)#	0.1(011)#			
S6	0.05(0111)#				

Figure 2.6 Huffman Coding Combined Symbols Are In Higer Position.

value of N. If N is 3, then the combined symbol is moved 3 positions higher than it would otherwise be. The effect of N for the same example as given in section A, can be seen in Figure 2.7.

Symbol	P	P
S1	0.4	0.4
S2	0.2	*0.1
S3	0.2	0.2
S4	0.1	0.2
S5	0.05#	0.1
S6	0.05#	

Figure 2.7 First Reduction For N = 3.

Huffman coding originally places the combined symbol just below the (0.1) value, since the probability is (0.1). But when N is 3, it is placed between (0.4) and (0.2). If we set N = 0.0, it gives the original Huffman coding.

b. Parameter K

K is an integer value. The probability of the combined value is multipled by K then the result is used as the new probability. The combined symbol is placed in decreasing probability order with respect to this new probability. Obviously if we set K = 1, we have the original Huffman coding. The first reduction of the modification, when K = 4, can be seen in Figure 2.8.

Symbol	P	P
S1	0.4	0.4
S2	0.2	*0.4
S3	0.2	0.2
S4	0.1	0.2
S5	0.05#	0.1
S6	0.05#	

Figure 2.8 First Reduction For K = 4.

c. *Parameter E*

E is a real number which is added to the combined value probability. So, the combined symbol is placed in decreasing probability order, with respect to this new result probability. The first reduction of modification process for E = 0.2 is given in Figure 2.9. Final code words and the average code length and variance are given in Figure 2.10. If we set E = 0.0, the original Huffman coding is reached.

Symbol	P	P
S1	0.4	0.4
S2	0.2	*0.3
S3	0.2	0.2
S4	0.1	0.2
S5	0.05#	0.1
S6	0.05#	

Figure 2.9 First Reduction For E = 0.2.

After using parameters N, K, and E at every step, the sum of the probabilities is no longer equal to 1.0. This does not affect the coding process and will be shown in Section D.

2. The Most Effective Parameter.

In [Ref. 2] the author used the parameters N, K, and E one at a time. As a result, the parameter E was found to be the most robust parameter because it provided better codes than parameters N and K. In my research I will use parameter E, but neither N nor K.

Symbol	P	P	P	P	P
S1	0.4(00)	0.4(01)	*0.5(00)	*0.7(1)	*1.1(0)
S2	0.2(11)	*0.3(10)	0.4(01)	0.5(00)#	0.7(1)
S3	0.2(000)	0.2(11)	0.3(10)#	0.4(01)#	
S4	0.1(001)	0.2(000)#	0.2(11)#		
S5	0.05(100)#	0.1(001)#			
S6	0.05(101)#				

Figure 2.10 Modified Huffman Coding For E = 0.2.

3. Dropping The Less Frequent Symbols

The main idea of this research is to code the source alphabet in such a way that as a result we will have a short average length and small variance. The general variance formula is

$$V = \sum P_i (X_i - L)^2$$

where, X_i = code length of the ith symbol.

The code lengths which are further away from the mean average length cause a large variance,since the second term of the above formula is squared. So, as a simple idea, if we bring them close to the mean, we have a small variance. In Huffman and Modified Huffman coding, the reasons for the large variance are the less frequent symbols, which have long code lengths individually. If we leave them out of our variance calculations one step earlier and out of the coding process, we can achieve a smaller variance.

In summary, the strategy is to determine the symbol frequencies, drop the less frequent symbols and then code using modified Huffman coding.

This idea is explained in the following calculations for the same source alphabet given in section B. For the sake of example, let's assume that symbols which have a probability less than 0.1 will be considered the less frequent symbols. The probability (P = 0.1) will be called the threshold, or limit, probability. In our example we have two symbols, S5 and S6, which have probabilities less than 0.1. Their

probabilities are 0.05 and 0.05, respectively. In the first step we drop these two symbols. See Figure 2.11.

Symbol	P	Symbol	P
S1	0.4	S1	0.4
S2	0.2	S2	0.2
S3	0.2	S3	0.2
S4	0.1	S4	0.1
S5	0.05		
S6	0.05		

Figure 2.11 Dropping Process For P= 0.1.

The second step is coding the rest of the symbols. For comparing the different techniques, the source alphabet will be coded, first by using Huffman coding for E = 0.2. Figure 2.12 shows the Huffman coding after the dropping process and Figure 2.13 shows the modified Huffman coding (E = 0.2) after the dropping process.

Symbol	P	P	P
S1	0.4(1)	0.4(1)	*0.5(0)
S2	0.2(01)	*0.3(00)#	0.4(1)
S3	0.2(000)#	0.2(01)#	
S4	0.1(001)#		

Figure 2.12 Huffman Coding After Dropping for P=0.1.

The average code length and variance calculation results of each coding technique are given in Figure 2.14.

As shown in Figure 2.14, dropping the less frequent symbols not only gave us a smaller variance, but also gave us a shorter code length. For Huffman coding, the reduction in average length after the dropping process is 26% and the reduction in variance is 60.16%. For modified Huffman coding, E = 0.2, the reduction in average

Symbol	P	P	P
S1	0.4(00)	*0.5(1)	*0.8(0)
S2	0.2(01)	0.4(00)#	0.5(1)
S3	0.2(10)#	0.2(01)#	
S4	0.1(11)#		

Figure 2.13 Modified Huffman Coding for P=0.1 and E = 0.2.

	Huffman	Modified H. After Drop.	Modified H.	Modified H.C. After Drop.
Aver.	2.3	1.7	2.4	1.8
Var.	1.81	0.721	0.24	0.036

Figure 2.14 Results Of The Four Different Coding Techniques.

length is 25% and in variance is 85%. If we compare Huffman coding and modified Huffman coding after dropping process, the reduction in average length is 21.7% and in variance is 98%.

At this point one important question arises, "What is the effect of dropping the less frequent symbols on the meaning of the original message?" To maintain the meaning of the messages, we need to choose a threshold probability P, for a given source alphabet, in a way that dropping the less frequent symbols does not affect the message meaning.

In Chapters 2 and 3, an optimal threshold probability (P), will be derived for a given source alphabet. In addition to maintaining message meaning, this optimal threshold probability will give us smaller average length and variance.

D. NORMALIZATION

Statistically, it is required that the sum of the symbol probabilities in the given source alphabet must be 1.0. During the modification of Huffman coding for K and E, at every step in the reduction process, the sum of the probabilities is no longer 1.0. In Figure 2.8 the sum of the symbol probabilties is 1.3 and in Figure 2.9 it is 1.2.

In [Ref. 2: p.18], it is shown that during the modification of Huffman coding using E and K, the same code words would be obtained with or without normalization. Similiarly, after dropping the less frequent symbols, the sum of the probabilities is not equal to 1.0. In Figure 2.11, it is 0.9. Figure 2.15 shows the Huffman coding after the dropping process with normalization. For normalization each symbol probability is divided by 0.9. The code words, which are obtained in Figure 2.12 and Figure 2.15 are exactly the same. Since the same code words are reached, will not apply normalization.

Symbol	P	Normalized P	P	P
S1	0.4	0.44(1)	0.44(1)	*0.56(0)
S2	0.2	0.22(01)	*0.34(00)#	0.44(1)
S3	0.2	0.22(000)#	0.22(01)#	
S4	0.1	0.12(001)#		

Figure 2.15 Huffman Coding After Dropping With Normalization.

III. MODIFICATION OF HUFFMAN CODING FOR TURKISH ALPHABET

A. SYMBOL FREQUENCIES IN TURKISH ALPHABET

As mentioned in Chapter 1, Huffman coding is a minimum - redundancy code, which uses the frequencies of the source symbol alphabet. The frequencies of the symbols in the Turkish alphabet were calculated in [Ref. 1] and [Ref. 2] by using the article given in Appendix A. The frequencies and the symbol probabilities of the Turkish alphabet are given in Table 1 and Table 2, respectively.

These frequencies approximate the real Turkish alphabet frequencies. The main difference is that some letters which are particular to the Turkish alphabet are not on the keyboard. During experiments these letters are represented by other Turkish letters in a way that the entire script can be understood by a Turkish reader with its original meaning. These particular letters and their representations are given in Figure 3.1.

Particular Letter	Representative
Ç	C
Ğ	G
İ	I
Ö	O
Ş	S
Ü	U

Figure 3.1 Particular Turkish Letters And The Representatives.

TABLE 1

SYMBOL CHARACTERISTICS OF THE TURKISH ALPHABET

Symbol	Frequecy	Cum. Freq.	Percent	Cum.Percent
.	182	182	1.017	1.017
(	12	194	0.067	1.084
)	15	209	0.084	1.168
;	11	220	0.061	1.229
-	3	223	0.017	1.246
space	2387	2610	13.339	14.585
,	219	2829	1.224	15.809
?	1	2830	0.006	15.814
:	6	2836	0.034	15.848
'	29	2865	0.162	16.010
"	20	2885	0.112	16.122
A	1687	4572	9.427	25.549
B	337	4909	1.883	27.432
C	293	5202	1.637	29.070
D	628	5830	3.509	32.579
E	1423	7253	7.952	40.531
F	64	7317	0.358	40.889
G	391	7708	2.185	43.073
H	104	7812	0.581	43.655
I	1884	9696	10.528	54.183
J	8	9704	0.045	54.227
K	691	10395	3.861	58.089
L	918	11313	5.130	63.219
M	527	11840	2.945	66.164
N	1183	13023	6.611	72.775
O	476	13499	2.660	75.434
P	123	13622	0.687	76.122
R	1089	14711	6.085	82.207
S	713	15424	3.984	86.192
T	575	15999	3.213	89.405
U	924	16923	5.163	94.568
V	156	17079	0.872	95.440
W	7	17086	0.039	95.479
X	1	17087	0.006	95.485
Y	480	17567	2.682	98.167
Z	177	17774	0.989	99.156
0	35	17779	0.196	99.352
1	24	17803	0.134	99.486
2	16	17819	0.089	99.575
3	13	17832	0.073	99.648
4	12	17844	0.067	99.715
5	15	17859	0.084	99.799
6	8	17867	0.045	99.844
7	5	17872	0.028	99.871
8	13	17885	0.073	99.944
9	10	17895	0.056	100.000

TABLE 2

SYMBOL PROBABILITIES

Symbol	Probability	Symbol	Probability
Space	0.13339	F	0.00358
I	0.10528	0	0.00196
A	0.09427	'	0.00162
E	0.07952	1	0.00134
N	0.06611	"	0.00112
R	0.06085	2	0.00089
U	0.05163	)	0.00084
L	0.05130	5	0.00084
S	0.03984	3	0.00073
K	0.03861	8	0.00073
D	0.03509	(	0.00067
T	0.03213	4	0.00067
M	0.02945	;	0.00061
Y	0.02682	9	0.00056
O	0.02660	J	0.00045
G	0.02185	6	0.00045
B	0.01883	W	0.00039
C	0.01637	:	0.00034
,	0.01224	7	0.00028
.	0.01017	-	0.00017
Z	0.00989	?	0.00006
V	0.00872	X	0.00006
P	0.00687	Q	0.00000
H	0.00581		

B. ASSIGNMENT OF THE CODES BY USING MODIFIED HUFFMAN CODING

Code representations are assigned to the Turkish alphabet, in the same procedure explained in section 1.A. Because of the length of the source alphabet, a computer program written in (LISP) language is used (Appendix B) [Ref. 1.] This program was run 50 times with the different values of the modification parameter E, from 0.0005 to 10.00. The output of the program are code words for each source symbol, average code length and variance. For E > 0.3, all the results show constant average length (5.08843) and constant variance (0.08061). Table 3 gives the average lengths and variances for each E value.

Generally a small increase in average length can give us a large reduction in variance. In other words, for smaller variances, there is a tradeoff in larger average length. The smallest variance (0.08061) corresponds the largest average length (5.08843) and vice versa. Some results demonsrated different behaviors. For example, average

length and variance are both increased (E = 0.0005 , average length = 4.30858, variance = 1.92289), but, the general tendency is clear. The trade - offs between average length and the variance are given in Figure 3.2. Table 4 gives the E values, average length and variance, in increasing order for average length. Table 5 also gives the same values in increasing order for variance.

During the experiments, one point is observed. The value of E is very effective up to a limiting value, but for values greater than this limiting value, the effect of E decreases. Figure 3.3 compares E and average length. If this figure is examined, it can be seen that up to E = 0.30, the increase in average length becomes constant. The same observation can be made in Figure. 3.4. In this figure, as E increases, the variance decreases. For E > 3.0, the variance also becomes constant. Figure 3.5 shows the effect of E on the average length and variance.

In Figure 3.2 the points closest to both axes are the extreme points and their values (E, average length and variance) are given in Figure 3.6. The graphic of these points, average length versus variance, is given in Figure 3.7. The code words which belong to these selected extreme codes are given in Table 6.

As mentioned, a gain (decrease) in the variance could be the result of a loss (increase) in average length. Table 7 gives the loss in average length and gain in variance for each experimental code. In the third column, negative variance gain indicates a loss. This is an exceptional case. Figure 3.8 shows the increase in the variance gain while the average length loss increases.

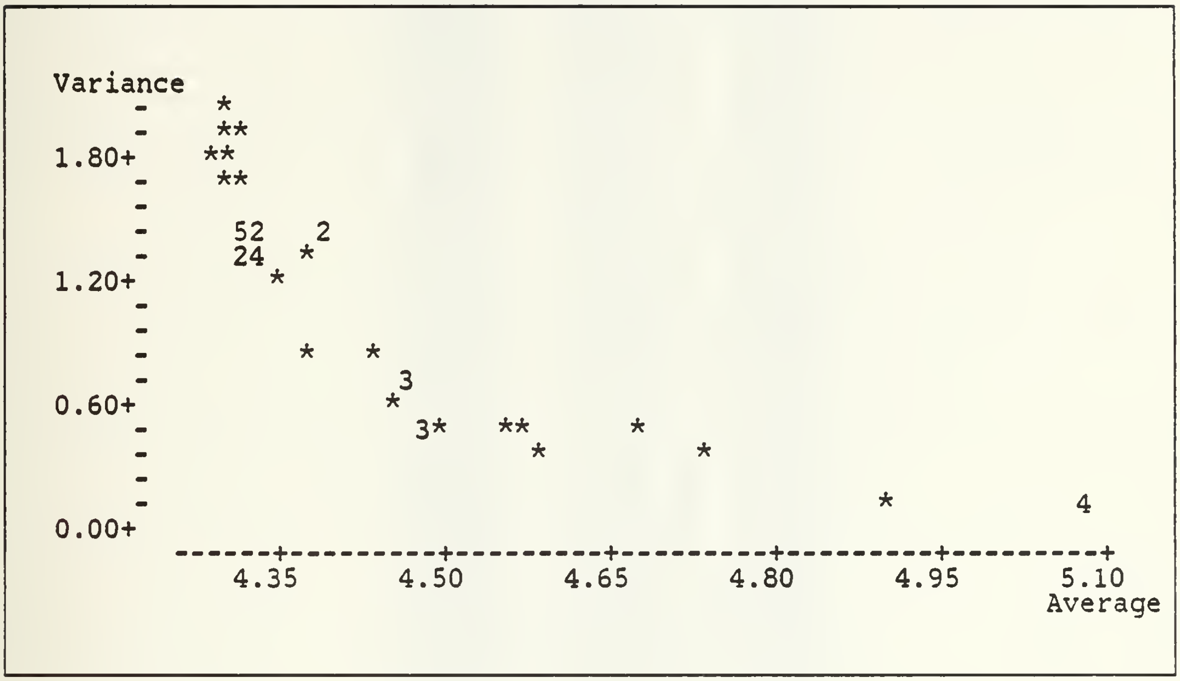

Figure 3.2 Average Length vs. Variance.

TABLE 3

RESULTS IN INCREASING E VALUE ORDER

E	Average	Variance
0.0	4.30771	1.91820
0.0005	4.30858	1.99289
0.0010	4.31181	1.74548
0.0015	4.31154	1.97072
0.0020	4.31371	1.42055
0.0025	4.31394	1.39718
0.0030	4.31293	1.93224
0.0035	4.31727	1.44681
0.0040	4.31583	1.43246
0.0045	4.31199	1.73289
0.0050	4.31961	1.73177
0.0055	4.32575	1.37092
0.0060	4.32217	1.42507
0.0065	4.32397	1.75761
0.0070	4.32046	1.34321
0.0075	4.33357	1.36628
0.0080	4.32759	1.35709
0.0085	4.33118	1.35962
0.0090	4.33145	1.35891
0.0095	4.34194	1.39358
0.0100	4.34066	1.38351
0.0150	4.36739	1.24489
0.0200	4.37334	1.35522
0.0250	4.39608	1.38716
0.0300	4.39537	1.38975
0.0350	4.47384	0.76056
0.0400	4.38381	0.89210
0.0450	4.47201	0.76166
0.0500	4.47085	0.75779
0.0550	4.44680	0.82955
0.0600	4.56179	0.48224
0.0650	4.46935	0.54266
0.0700	4.48795	0.51705
0.0750	4.59420	0.41606
0.0800	4.49995	0.50834
0.0850	4.48231	0.50933
0.0900	4.48231	0.50933
0.1000	4.57556	0.42755
0.1500	4.73389	0.34298
0.2000	4.68298	0.42142
0.2500	4.89779	0.16814
0.3000	5.08843	0.08061
0.3500	5.08843	0.08061
0.4000	5.08843	0.08061
0.5000	5.08843	0.08061

TABLE 4

RESULTS IN INCREASING AVERAGE LENGTH ORDER

Average	E	Variance
4.30771	0.00	1.91820
4.30858	0.0005	1.99289
4.31154	0.0015	1.97072
4.31181	0.0010	1.74548
4.31199	0.0045	1.73289
4.31293	0.0030	1.93224
4.31371	0.0020	1.42055
4.31394	0.0025	1.39718
4.31583	0.0040	1.43246
4.31727	0.0035	1.44681
4.31961	0.0050	1.73177
4.32046	0.0070	1.34321
4.32217	0.0060	1.42507
4.32397	0.0065	1.75761
4.32575	0.0055	1.37092
4.32759	0.0080	1.35709
4.33118	0.0085	1.35962
4.33145	0.0090	1.35891
4.33357	0.0075	1.36628
4.34066	0.0100	1.38351
4.34194	0.0095	1.39358
4.36739	0.0150	1.24489
4.37334	0.0200	1.35522
4.38381	0.0400	0.89210
4.39537	0.0300	1.38975
4.39608	0.0250	1.38716
4.44680	0.0550	0.82955
4.46935	0.0650	0.54266
4.47085	0.0500	0.75779
4.47201	0.0450	0.76166
4.47384	0.0350	0.76056
4.48231	0.0850	0.50933
4.48231	0.0900	0.50933
4.48795	0.0700	0.51705
4.49995	0.0800	0.50834
4.56179	0.0600	0.48224
4.57556	0.1000	0.42755
4.59420	0.0750	0.41606
4.68298	0.2000	0.42142
4.73389	0.1500	0.34298
4.89779	0.2500	0.16814
5.08843	0.3000	0.08061
5.08843	0.3500	0.08061
5.08843	0.4000	0.08061
5.08843	0.5000	0.08061

TABLE 5

RESULTS IN INCREASING VARIANCE ORDER

Variance	E	Average
0.08061	0.5000	5.08843
0.08061	0.4000	5.08843
0.08061	0.3500	5.08843
0.08061	0.3000	5.08843
0.16814	0.2500	4.89779
0.34298	0.1500	4.73389
0.41606	0.0750	4.59420
0.42142	0.2000	4.68298
0.42755	0.1000	4.57556
0.48224	0.0600	4.56179
0.50834	0.0800	4.49995
0.50933	0.0850	4.48231
0.50933	0.0900	4.48231
0.51705	0.0700	4.48795
0.54266	0.0650	4.46935
0.75779	0.0500	4.47085
0.76056	0.0350	4.47384
0.76166	0.0450	4.47201
0.82955	0.0550	4.44680
0.89210	0.0400	4.38381
1.24489	0.0150	4.36739
1.34321	0.0070	4.32046
1.35522	0.0200	4.37334
1.35709	0.0080	4.32759
1.35891	0.0090	4.33145
1.35962	0.0085	4.33118
1.36628	0.0075	4.33357
1.37092	0.0055	4.32575
1.38351	0.0100	4.34066
1.38716	0.0250	4.39608
1.38975	0.0300	4.39537
1.39358	0.0095	4.34194
1.39718	0.0025	4.31394
1.42055	0.0020	4.31371
1.42507	0.0060	4.32217
1.43246	0.0040	4.31583
1.44681	0.0035	4.31727
1.73177	0.0050	4.31961
1.73289	0.0045	4.31199
1.74548	0.0010	4.31181
1.75761	0.0065	4.32397
1.91820	0.0	4.30771
1.93224	0.0030	4.31293
1.97072	0.0015	4.31154
1.99289	0.0005	4.30858

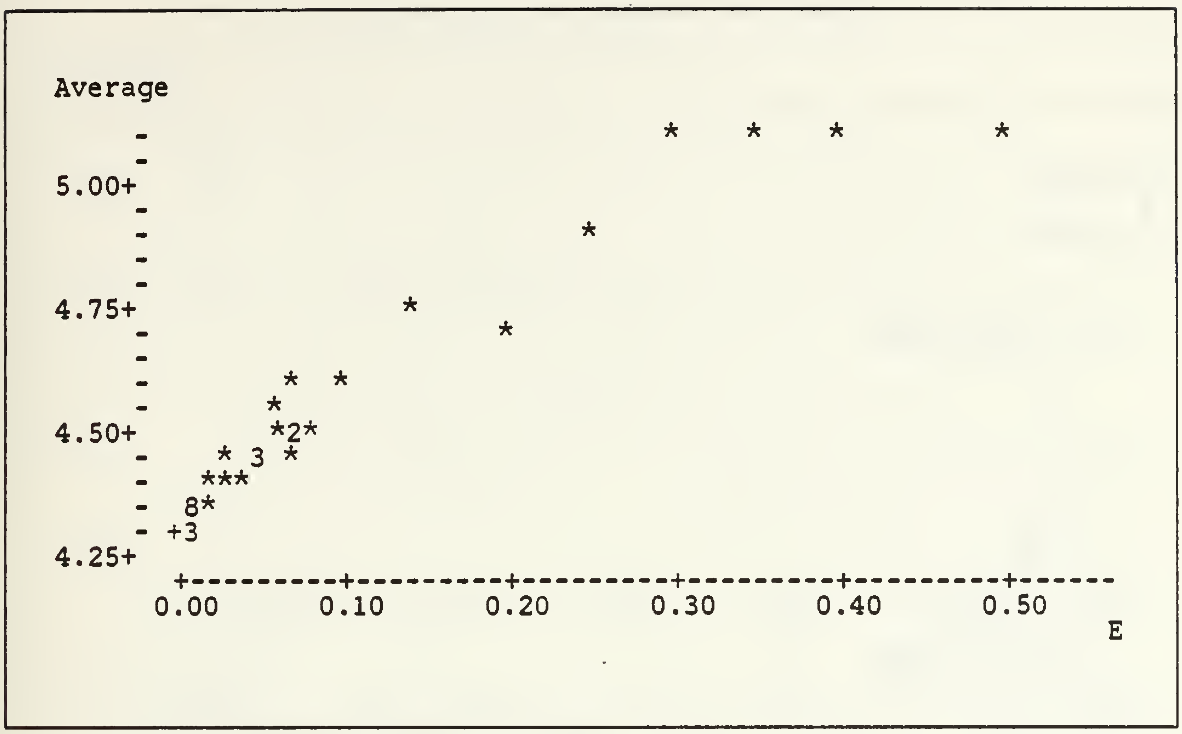

Figure 3.3 Effect of the parameter E on average length.

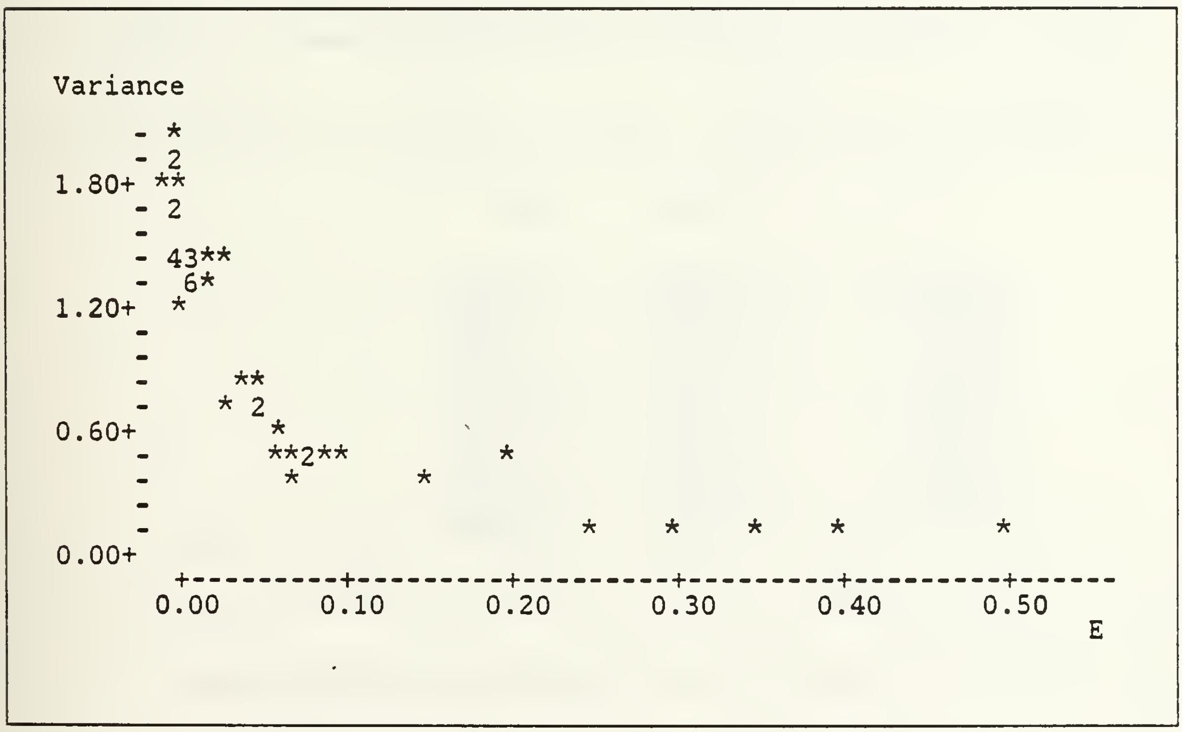

Figure 3.4 Variance vs. Parameter E.

```
Average
&
Variance
      -                                        A     A     A           A
      -                   A     A     A
  4.5+  A22A3A32AA
      - ++
      -
      -
      -
  3.0+
      -
      -
      - 2
      - 4B
  1.5+ 482B
      -  2
      -    BB3
      -       B22B
      -        B  B     B     B     B
  0.0+                                     B     B     B           B
        +---------+---------+---------+---------+---------+------
       0.00      0.10      0.20      0.30      0.40      0.50
                                                              E

          A = Ave vs. E                B = Var vs. E
```

Figure 3.5 Effect of E on Average And variance.

E	AVER.	VAR.
0.0000	4.30771	1.91820
0.0005	4.30858	1.92289
0.0010	4.31181	1.74548
0.0045	4.31199	1.73289
0.0055	4.32575	1.37092
0.0400	4.38381	0.89210
0.0800	4.49995	0.50834
0.0750	4.59420	0.41606
0.1500	4.73389	0.34298
0.2500	4.89779	0.16814
0.3000	5.08843	0.08061

Figure 3.6 Results And The E Value Of The Experimental Codes.

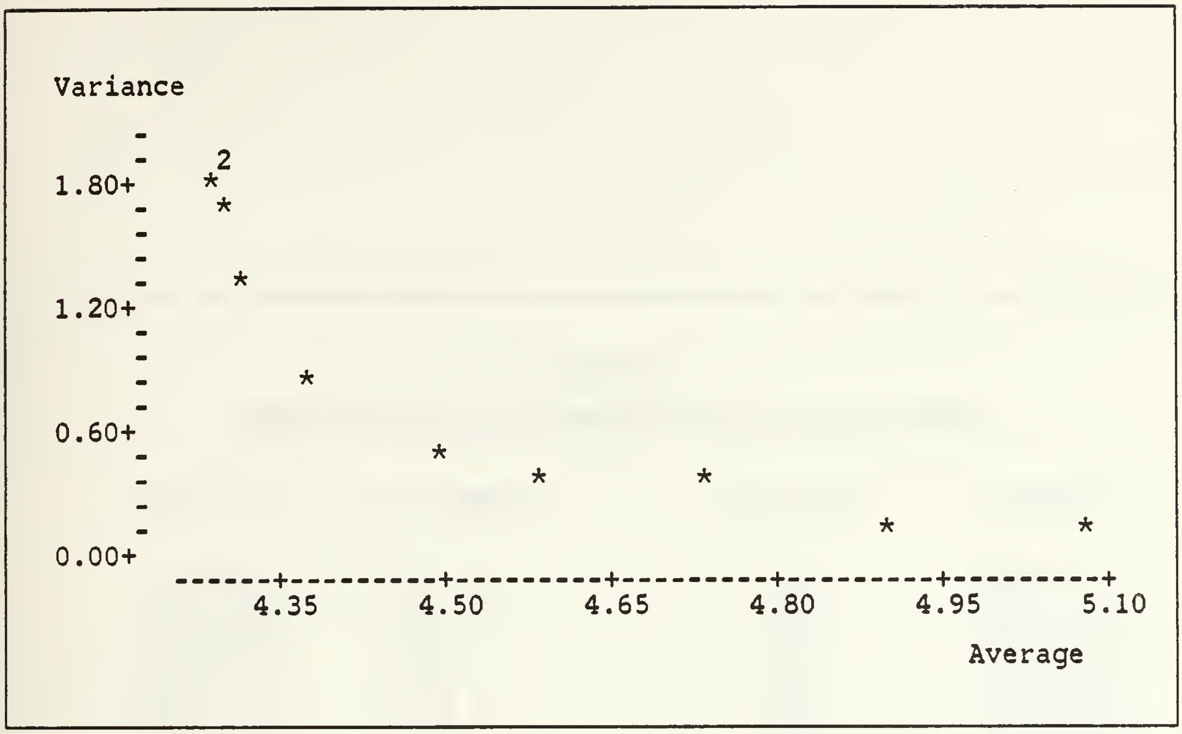

Figure 3.7 Average Lengths vs. Variance.

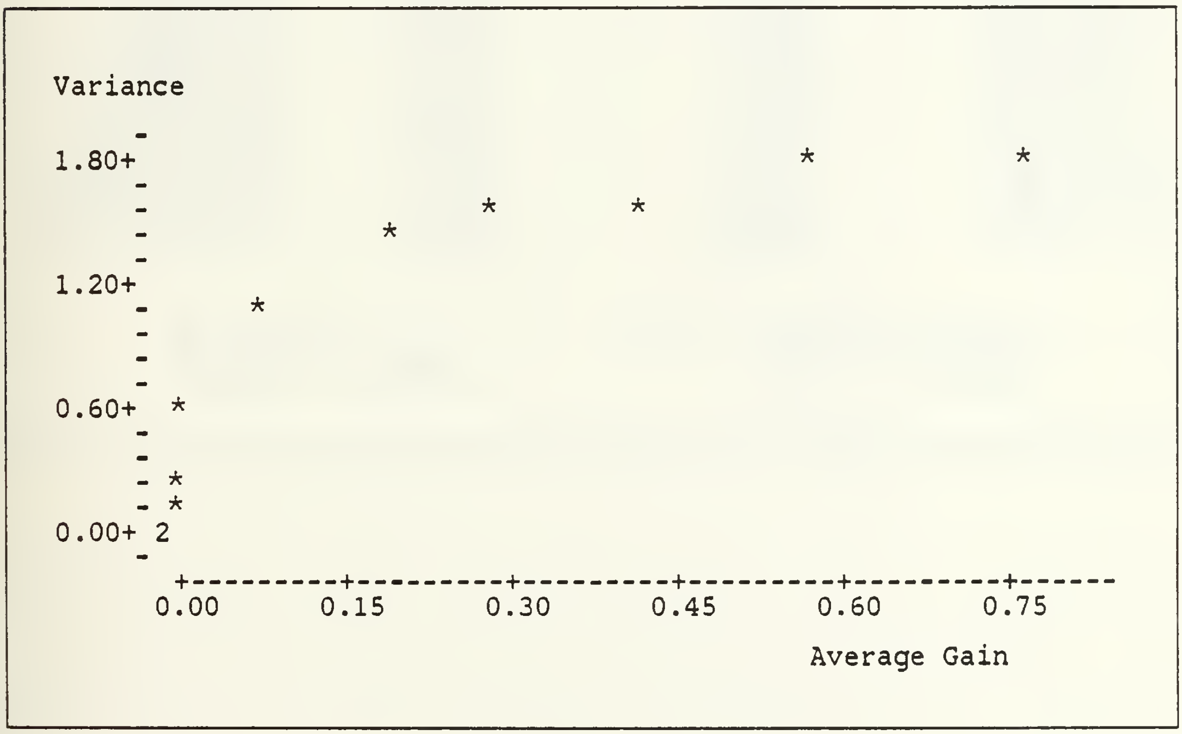

Figure 3.8 Loss In Average Length vs. Gain In Variance.

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	010	F	10010011
I	101	0	100100100
A	111	'	100101101
E	0001	1	0000011100
N	0110	"	0000011101
R	1000	2	1001001011
U	1100	)	1001011001
L	1101	5	1001011000
S	00101	3	1001011110
K	00101	8	1001011101
D	00111	(	1001011111
T	01110	4	00000111010
M	01111	;	00000111011
Y	10011	9	00000111101
O	000000	J	10010010101
G	000010	6	10010010100
B	001100	W	10010111000
C	001101	:	10010111001
,	0000010	7	000001111001
.	0000110	-	0000011110000
Z	0000111	?	0000011110001 1
V	1001000	X	00000111100010 0
P	1001010	Q	00000111100010 1
H	00000110		

E = 0.0 (Huffman Coding)
Average = 4.30771
Variance = 1.91820

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	010	F	10111110
I	101	0	000111110
A	111	'	001100000
E	1000	1	0001011110
N	0110	"	1101011110
R	0001	2	0011011110
U	0011	)	1011011110
L	1011	5	1111011110
S	00100	3	0111011110
K	10100	8	1100111110
D	11100	(	0100111110
T	01110	4	01011100000
M	01000	;	01001101111
Y	11001	9	00111100000
O	000000	J	01111100000
G	010000	6	10111100000
B	001100	W	11111100000
C	101100	:	00101011110
,	0100000	7	10101011110
.	0110000	-	011011100000
Z	1110000	?	011001011110
V	1110000	X	111011100000
P	1111110	Q	111001011110
H	01100000		

E = 0.0005
Average = 4.30858
Variance = 1.92290

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	010	F	00111110
I	101	0	010011011
A	111	'	1001101100
E	1000	1	0101101100
N	0110	"	0011101100
R	0001	2	0111101100
U	0011	)	0000011011
L	00000	5	1111101100
S	00100	3	0100011011
K	10100	8	1000011011
D	11100	(	1100011011
T	01110	4	0010111110
M	01001	;	1010111110
Y	11001	9	0110111110
O	01011	J	0110011011
G	010000	6	1110111110
B	001100	W	1110011011
C	011110	:	00001101100
,	111011	7	10001101100
.	0110000	-	01101101100
Z	1110000	?	01011101100
V	0101100	X	11101101100
P	1111110	Q	11011101100
H	1011011		

E = 0.001
Average = 4.31181
Variance = 1.74548

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	101	F	11001000
I	1000	0	00011100
A	1010	'	10011100
E	1110	1	01011100
N	0001	"	11011100
R	1001	2	00111100
U	1111	)	01111100
L	01100	5	10111100
S	10011	3	11111100
K	00000	8	011000100
D	10000	(	111000100
T	00010	4	0000000100
M	10010	;	0100000100
Y	01011	9	1000000100
O	11011	J	1100000100
G	00110	6	0010000100
B	10110	W	0110000100
C	00111	:	1010000100
,	10111	7	1110000100
.	010100	-	0001000100
Z	110100	?	1001000100
V	000011	X	0101000100
P	100011	Q	1101000100
H	0100100		

E = 0.04
Average = 4.38381
Variance = 0.89202

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	110	F	00010011
I	101	0	10010011
A	111	'	101101000
E	0100	1	011101000
N	1010	"	111101000
R	0001	2	000010000
U	1011	)	100010000
L	00000	5	0010101000
S	11000	3	0110101000
K	11100	8	1010101000
D	00010	(	1110101000
T	10010	4	0001101000
M	01001	;	1001101000
Y	11001	9	01000101000
O	00011	J	11000101000
G	001000	6	00100101000
B	001100	W	01100101000
C	101100	:	10100101000
,	110011	7	11100101000
.	1010000	-	000000101000
Z	0110000	?	100000101000
V	1110000	X	010000101000
P	1010011	Q	110000101000
H	10010000		

E = 0.0045
Average = 4.31199
Variance = 1.73289

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	110	F	01101111
I	011	0	11101111
A	0000	'	100101000
E	0100	1	010101000
N	1010	"	110101000
R	0001	2	010111001
U	1101	)	110111001
L	0111	5	001111001
S	11000	3	011111001
K	01100	8	101111001
D	00010	(	111111001
T	10010	4	001101000
M	01001	;	001111100
Y	00101	9	101101000
O	10101	J	101111100
G	001000	6	011111100
B	011100	W	111111100
C	011001	:	0000111001
,	001111	7	1000111001
.	011111	-	0100111001
Z	0111100	?	1100111001
V	0111100	X	0000101000
P	0101111	Q	1000101000
H	11101000		

E = 0.0055
Average = 4.32575
Variance = 1.37092

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	0010	F	101001
I	0001	0	011001
A	0101	'	111001
E	0011	1	0011101
N	1111	"	1011101
R	00000	2	0111101
U	10000	)	1111101
L	01000	5	0001101
S	11000	3	0101101
K	00100	8	1001101
D	10100	(	1101101
T	01100	4	0011101
M	11100	;	1011011
Y	00110	9	0111011
O	10110	J	1111011
G	01110	6	00001101
B	11110	W	01001101
C	00111	:	10001101
,	10111	7	00101101
.	001010	-	11001101
Z	101010	?	01101101
V	011010	X	10101101
P	111010	Q	11101101
H	001001		

E = 0.075
Average = 4.5942
Variance = 0.41607

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	0010	F	1001011
I	0001	0	0101001
A	0101	'	1101001
E	0011	1	0011001
N	1011	"	1011001
R	0111	2	0111001
U	1111	)	1111001
L	00000	5	00001101
S	10000	3	01001101
K	01000	8	10001101
D	11000	(	00101101
T	00100	4	11001101
M	10100	;	10101101
Y	01100	9	01101101
O	11100	J	00011101
G	00110	6	11101101
B	10110	W	10011101
C	01110	:	01011101
,	11110	7	11011101
.	001010	-	00111101
Z	101010	?	01111101
V	011010	X	10111101
P	111010	Q	11111101
H	0001001		

E = 0.08
Average = 4.49995
Variance = 0.50838

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	1101	F	101000
I	0011	0	011000
A	1011	'	111000
E	01100	1	000100
N	11100	"	100100
R	00010	2	010100
U	10010	)	110100
L	01010	5	0000111
S	11010	3	1000111
K	00110	8	0100111
D	10110	(	1100111
T	01110	4	0010111
M	11110	;	0110111
Y	00001	9	1010111
O	10001	J	0001111
G	01001	6	1110111
B	11001	W	0101111
C	00101	:	1001111
,	10101	7	1101111
.	000000	-	0011111
Z	100000	?	0111111
V	010000	X	1011111
P	110000	Q	1111111
H	001000		

E = 0.15
Average = 4.73389
Variance = 0.34298

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	1111	F	010010
I	00000	0	100010
A	10000	'	110010
E	01000	1	001010
N	11000	"	011010
R	00110	2	101010
U	10110	)	111010
L	01110	5	0000100
S	11110	3	1000100
K	00001	8	0100100
D	10001	(	1100100
T	01001	4	0010100
M	11001	;	1010100
Y	00101	9	0110100
O	10101	J	1110100
G	01101	6	0001100
B	11101	W	1001100
C	00011	:	0101100
,	10011	7	1101100
.	01011	-	0011100
Z	00111	?	0111100
V	11011	X	1011100
P	10111	Q	1111100
H	000010		

E = 0.25
Average = 4.89779
Variance = 0.16814

TABLE 6

CODE WORDS OF THE EXPERIMENTAL CODES

Symbol	Code Word	Symbol	Code Word
Space	11100	F	111000
I	00010	0	000110
A	10010	'	100110
E	01010	1	010110
N	11010	"	110110
R	00101	2	001110
U	10101	)	011110
L	01101	5	101110
S	11101	3	000001
K	00011	8	111110
D	10011	(	010001
T	01011	4	100001
M	11011	;	110001
Y	00111	9	001001
O	10111	J	011001
G	01111	6	101001
B	11111	W	111001
C	000000	:	000100
,	100000	7	100100
.	010000	-	010100
Z	110000	?	001100
V	001000	X	110100
P	101000	Q	101100
H	011000		

E = 0.30
Average = 5.08843
Variance = 0.08061

TABLE 7

LOSS IN AVERAGE LENGTH AND GAIN IN VARIANCE

E	Average Lost	Variance Gain
0.0000	0.000000	0.00000
0.0005	0.000870	-0.00469
0.0010	0.004100	0.17272
0.0045	0.004280	0.18531
0.0055	0.018040	0.54728
0.0400	0.076099	1.02610
0.0800	0.192240	1.40986
0.0750	0.286489	1.50214
0.1500	0.426180	1.57522
0.2500	0.590080	1.75006
0.3000	0.780720	1.83759

C. MODIFIED HUFFMAN CODING AFTER DROPPING THE LESS FREQUENT SYMBOLS

In dropping the less frequent source symbols, the main idea is to set the limit probability = P. The symbols which have a lower probability value than the limit probability are dropped. If the P value is very high, the meaning of the message might be disturbed. On the contrary, if the P value is very small, the dropping process will have little or no effect on the average code length and variance.

In this section, I examined seven different P values. At each step, I dropped the symbols with probabilities lower than P and ran the same LISP program for the experimental E values given in Figure 3.6.

One point must be mentioned. In every type of message the numbers have a very important place. Hence, when the numbers are represented numerically, even if they have a lower probability than the limit probability, they are not dropped.

At each step, the effect of dropping the source symobls on the meaning of the message is the subject of the next chapter. Here I examined the technical aspect. In other words, disregarding the meaning of the information, I increased the P value and examined the changes of the average code length and variance for experimental E values.

The limit probabilities (P) were chosen arbitrarily. These limit probabilities and corresponding step numbers are given in Figure 3.9.

At every step symbols which have lower probabilities than the P value are dropped. Table 8 shows the dropped symbols in each step.

The results were examined in two dimensions. In the first, changes in average length and variance for ecah P value were examined, while using the experimental E values. In the second, changes in average length and variance for each E value, while using selected P values were examined. All results are given in Table 9. For each E value and step, the average length and the variance can be seen.

1. Evaluation of the First Dimension

As mentioned earlier, the first dimension is the behavior of the average length and variance for each P value, while employing experimental E values. The purpose is to understand the results as the P value is increased.

This dimension is represented in Table 9, in rows, for each E value. The last row consists of mean average lengths and mean variances.

Limit Probabilities	Step Number
0.0004	1
0.0009	2
0.000175	3
0.0006	4
0.009	5
0.015	6
0.025	7

Figure 3.9 Limit Probabilities At Each Step.

TABLE 8

DROPPED SYMBOLS AT EACH STEP

Step Number	Dropped Symbol
1	q x ? - : w
2	Step 1 symbols and j ; ( )
3	Step 2 symbols and " '
4	Step 3 symbols and f h
5	Step 4 symbols and p v
6	Step 5 symbols and z . ,
7	Step 6 symbols and c b g

When the last row is examined, it can easily be seen that mean values tend to decrease as P increases. In other words, the more symbols that are dropped, the smaller average code length and variance reached. These last row values are given in Figure 3.10. The first row of Figure 3.10 gives the mean values without dropping any symbols. This is represented as step 0. Changes in the mean average length and mean variance, as the P value increases (each P corresponds to a step number) are given in Figure 3.11 and Figure 3.12, respectively.

TABLE 9

RESULTS OF EACH STEP FOR EXPERIMENTAL E VALUES

E	Step1	Step2	Step3	Step4	Step5	Step6	Step7	Mean
0.0	4.2985 1.8441	4.25824 1.72806	4.25871 1.76745	4.2088 1.52891	4.1408 1.3175	4.0135 0.9597	3.825 1.0069	4.163 1.450
0.0005	4.2985 1.8464	4.2786 1.71537	4.25887 1.76699	4.20932 1.53131	4.1413 1.3200	4.01413 0.96142	3.826 1.0089	4.150 1.451
0.001	4.2992 1.8464	4.27963 1.71537	4.25886 1.76699	4.20932 1.53131	4.1413 1.3200	4.01413 0.96142	3.826 1.0089	4.147 1.451
0.0045	4.3136 1.4017	4.28641 1.24142	4.27408 1.15705	4.21096 1.51672	4.1431 1.2720	4.01413 0.96142	3.8260 1.0089	4.152 1.223
0.0055	4.3092 1.3282	4.27246 1.92706	4.27246 1.15441	4.22498 1.31839	4.1431 1.2709	4.01413 0.96142	3.8327 0.6951	4.156 1.132
0.0400	4.3751 0.8299	4.36127 0.76136	4.34814 0.71132	4.35133 0.60011	4.2715 0.5438	4.07948 0.48310	3.9243 0.5410	4.244 0.638
0.0800	4.4912 0.4708	4.40419 0.60980	4.38820 0.52578	4.36230 0.36693	4.3152 0.2803	4.24123 0.21604	3.8904 0.4923	4.299 0.423
0.0750	4.5616 0.3736	4.56811 0.36446	4.57221 0.35245	4.41920 0.31971	4.4192 0.3023	4.24135 0.20952	4.1263 0.1366	4.425 0.294
0.1500	4.4903 0.5091	4.47942 0.46734	4.68289 0.25246	4.42917 0.31476	4.6583 0.2289	4.31951 0.22842	4.1463 0.1366	4.458 0.305
0.2500	5.0232 0.0226	4.70216 0.28173	4.87094 0.12277	4.66697 0.23258	4.5112 0.2605	4.31960 0.53337	4.3278 0.2204	4.631 0.239
0.3000	5.0232 0.0226	5.00763 0.00758	4.86094 0.12274	4.66697 0.23258	4.3058 0.3571	4.55952 0.24646	4.3279 0.2204	4.693 0.173
Mean	4.4940 0.9542	4.45002 0.98360	4.45966 0.88176	4.36087 0.86303	4.2982 0.7704	4.16619 0.6112	4.08176 0.52545	

The general tendency is that the mean value decreases as the P value increases. On the other hand, some P values have an effect which is contrary to the

general tendency. For example, in step 3 (P = 0.000175) there is an increase in both mean values over those in step 2. Another one is that for E = 0.08000, in step 2 and 3 (P2 = 0.0009 and P3 = 0.000179) the variances are larger than the variance of step 1.

These experimental values are the results of the numbers′ nature in the reduction process of the Huffman coding. This is the reason why each source alphabet should be examined separately. Each has its own optimal P and E values.

Additionally, it should be mentioned that an optimal E value for a specific P value might not be optimal for another P value, and vice versa, the P value which is optimal for any E value might not be optimal for another E value. For example, in Table 9, in step 2 (P = 0.0009), for E = 0.00400, the variance is 1.22708, which is smaller than the variance of step 1 for the same E. But, for the same P value (step 2), for E = 0.08000, the variance is 0.60980, which is larger than the variance of step 1 for the same E.

Since the limit probability, P, has some effect on the meaning of the messages, a P value should first be chosen for a given source alphabet in a way that will not destroy the meaning of the messages. Then the optimal E value, which gives the optimal average length and variance in Huffman coding, should be chosen for the optimal P.

The fourth chapter of this study, after examining the effect of these seven experimental P values on the Turkish messages, finds the optimal P.

Step	Meanav.	Meanva.
0	4.95512	1.01542
1	4.49942	0.95417
2	4.45002	0.98360
3	4.45966	0.88176
4	4.36087	0.86303
5	4.29825	0.77041
6	4.16619	0.61112
7	4.08176	0.52545

Figure 3.10 Mean Average Lengths And Mean Variances.

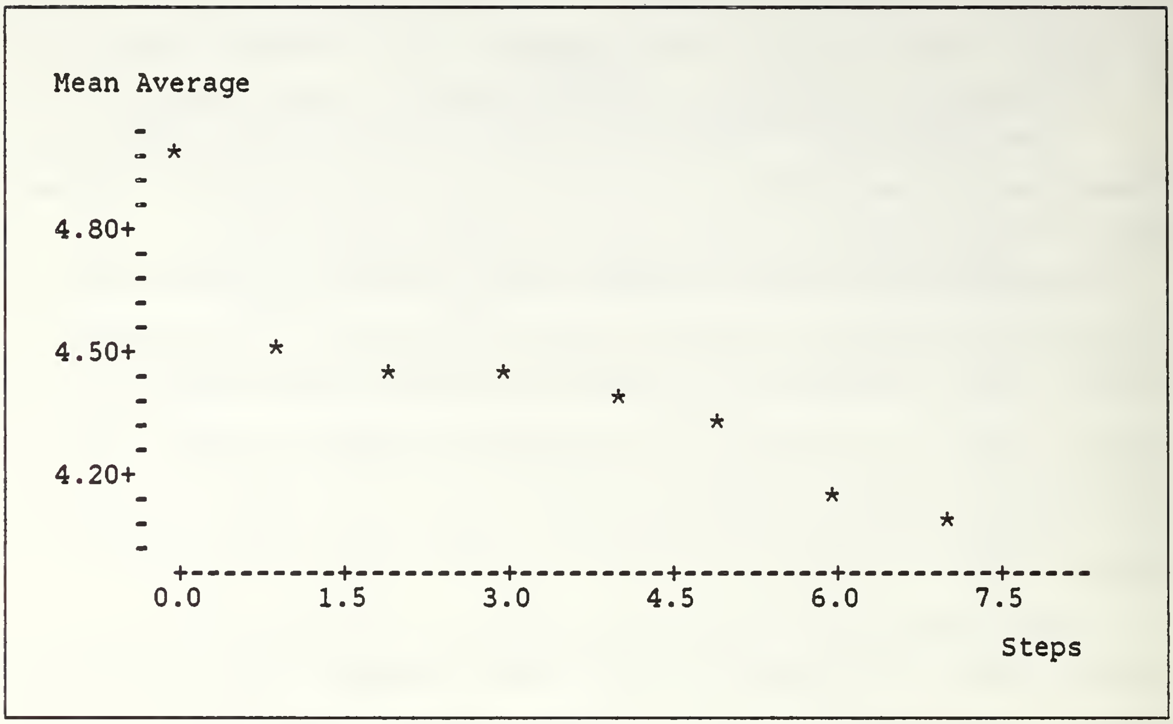

Figure 3.11 Mean average Length vs. Steps.

2. Evaluation of the Second Dimension

The second dimension of the results involves the changes in average length and variance for each E value, while using the different P values. This dimension is represented by columns for each P value in Table 9. It shows us the changes in average length and variance for each experimental E values while employing each P value. The last column of Table 9 gives the mean values for each E value, mean average lengths, and mean variances. By examining the last column, we can see the behavior of the average length and variance for each E value, with the total effect of different P values. These last column values are given in Figure 3.13. The changes in mean average length with respect to different E values and in mean variance with respect to different E values are given in Figure 3.14 and Figure 3.15, respectively.

The general tendency is for mean average length to increase as the E value increases and mean variance to decrease as the E value increases. For some exceptional values, the same comment can be made as was made previously. Thus, the optimal E value should be chosen for each P value separately.

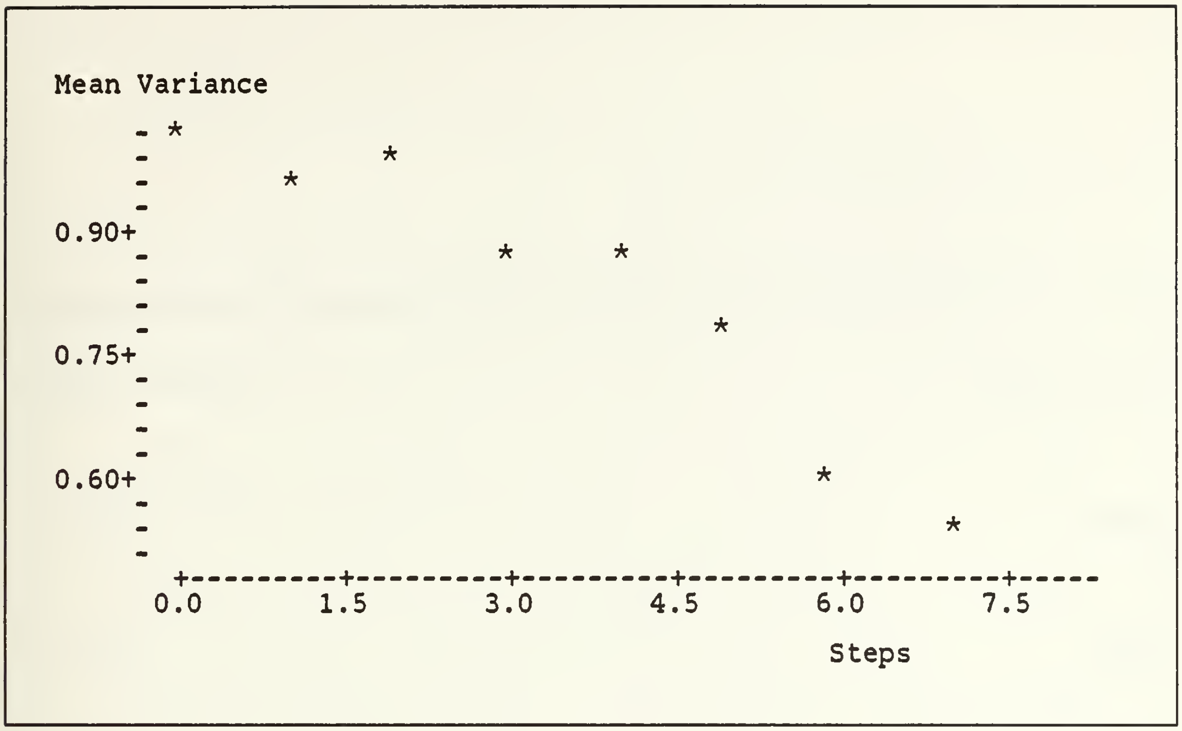

Figure 3.12 Mean Variance vs. Steps.

E	MEANAVE.	MEANVAR.
0.0000	4.14630	1.45000
0.0050	4.14670	1.45100
0.0010	4.14690	1.45000
0.0045	4.15260	1.22280
0.0055	4.15590	1.31180
0.0400	4.24444	0.63866
0.0800	4.29910	0.42410
0.0750	4.42260	0.29410
0.1500	4.45800	0.30540
0.2500	4.63100	0.23910
0.3000	4.69300	0.17280

Figure 3.13 Mean Average Lengths And Mean Variances For Each E.

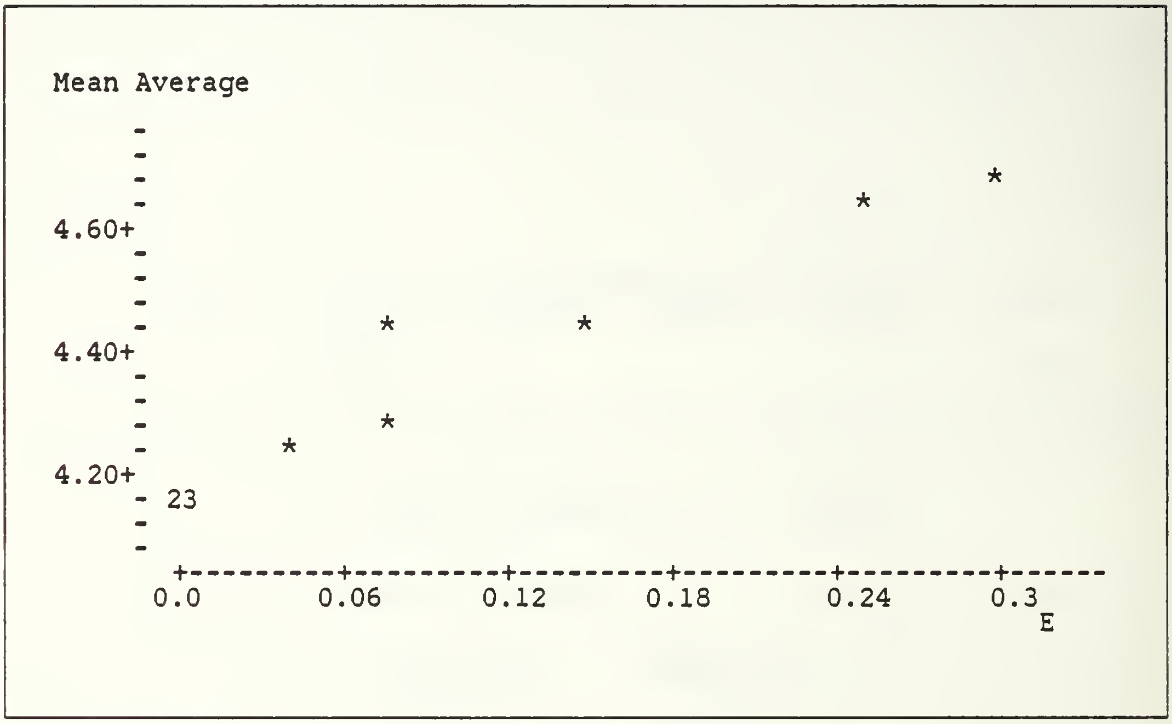

Figure 3.14 Mean Average vs. E Values.

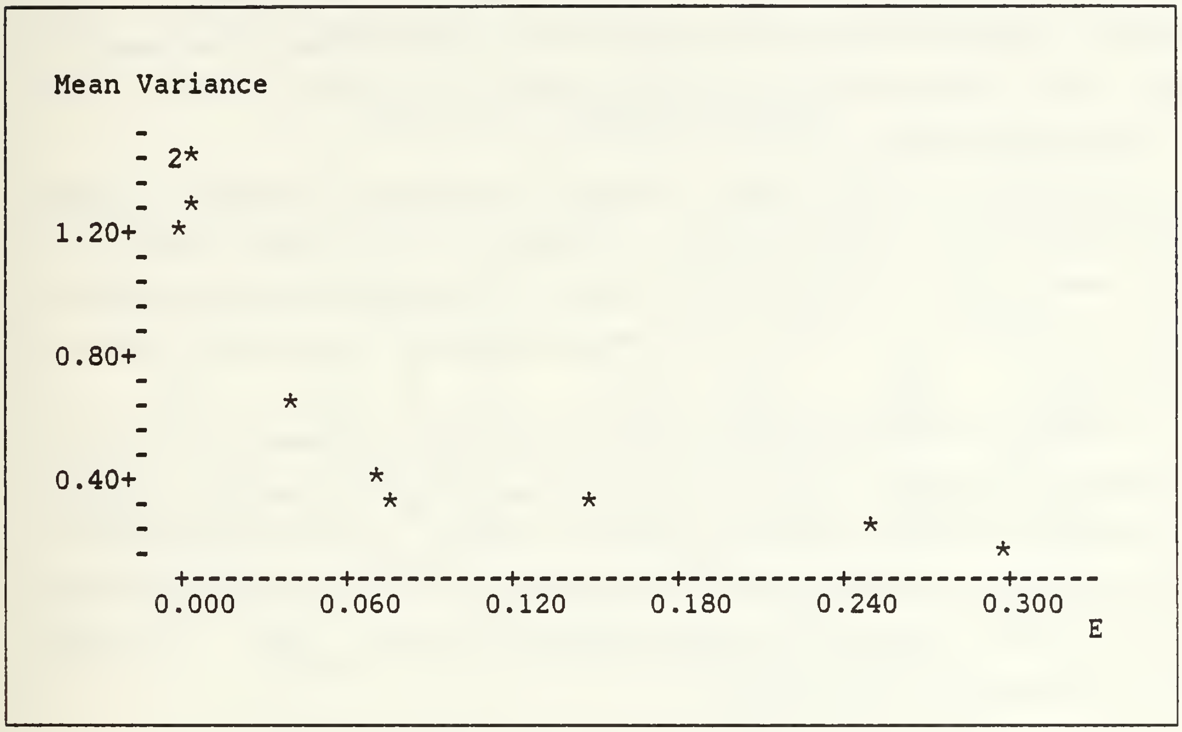

Figure 3.15 Mean Variance vs. E Value.

IV. STATISTICAL EVALUATION OF THE DROPPING PROCESS

A. THE DROPPING PROCESS

In the previous chapter, the technical evaluation of the dropping process was discussed. In seven steps, different P values for each step (see Figure 3.9) were employed. The results of these decoding processes for each P value showed that, as P increased, the average code length and variance decreased. But the results still contain the "average code length vs. variance" trade - off.

In this chapter, the main issue is to prevent the meaning of the message from any distortion during the dropping process. Although the more symbols dropped, the smaller average length and variance reached, in real life applications we cannot drop as many as we would like. At this point a limit probability (P) becomes the subject of discussion.

To find the limit probability (P) for the given Turkish alphabet, four different short articles are examined using the Pascal language computer program in Appendix C. Each article is rewritten seven times using this program. In each step, experimental P values, which were given in Figure 3.9 are employed. The Pascal program does not rewrite the symbols which have lower probabilities than the P value. The original short articles are given in Appendix D.

B. STATISTICAL EVALUATION FOR LIMIT PROBABILITY (P)

After each step, the four short articles were read by Turkish officers attending N.P.S., in order to grade the meaning level of each article. Fifteen officers graded these articles, with grade ranges from 0 to 4. 0 means nothing is understandable and 4 corresponds to the level that meaning is very clear. These grade numbers and corresponding meaning levels are given in Figure 4.1.

The results of the survey showed that the meaning level up to the seventh step exhibited a slow decrease. These decreases stayed in the clear level. But, in the seventh step, it suddenly dropped into the very difficult region. Figure 4.2 gives the average grades for each article at every step. Figure 4.3 shows the resulting average meaning level of each step. The change in the meaning level wiile the limit probability (P) increases can be examined in Figure 4.4.

Grade Number	Meaning Level
0	Nothing is understood.
1	Very diffucult.
2	Difficult.
3	Meaning is clear.
4	Meaning is very clear.

Figure 4.1 Grade Classification.

Article No	S1	S2	S3	S4	S5	S6	S7
1	4	4	4	4	4	3.72	2.35
2	4	4	4	3.95	3.82	3.43	2.17
3	4	4	4	3.13	3.09	2.82	1.65
4	4	4	4	4	3.91	3.10	1.70
Total Grade:	16	16	16	15.08	14.82	13.07	7.88

Figure 4.2 Grades For Each Article At Every Step.

Step Number	1	2	3	4	5	6	7
Meaning Level	4	4	4	3.77	3.70	3.26	1.97

Figure 4.3 Average Meaning Levels At Each Step.

C. OPTIMAL LIMIT PROBABILITY

Our purpose is to choose the optimal probability for the given Turkish alphabet. This optimal limit probability is supposed to give a decrease in average length and variance, while remaining in the clear level. This logic leads us, by examining Figure 4.4, to choose step 6 probability as the optimal one. The optimal limit probablity,

used in step 6, is 0.015. The rewritten forms of the articles at step 6 are given in Appendix E.

In Chapter 3, Table 9 gives the average lengths and variances for each step (P value) and for each E value. The average lengths and variance for P = 0.015 and corresponding E values are given in Figure 4.5. The code words which are the results of the experimental E values and P = 0.015 are given in Table 10.

The trade - offs between average length and variance, after the dropping process, are given in Figure 4.6. The same conclusion found in Chapter 2 could be reached, namely a decrease in variance requires an increase in the average length.

The modified Huffman coding results, average length and variance values for experimental E values, without dropping any symbol (given in Figure 3.6) and the values after dropping the symbols which have lower probabilities than 0.015 are compared.

The average length and variance differences between the encoding processes, before and after dropping the symbols for P = 0.015, are given in Figure 4.7. In this figure, the positive values show the decreases and the negative values show the increases in the results after dropping.

Generally, a decrease can be seen in both average length and in variance. But for E values of 0.2500 and 0.3000, the variances after the dropping process increased, by 0.36523 and 0.16585, respectively. Figure 4.8 shows the change in average lengths before and after dropping while E increases. It can be seen from Figure 4.9 that, as the E value increases, the difference between average lengths, before and after dropping, also increases. Further, E = 0.25 and larger values, the variance after dropping becomes larger than the variance before dropping.

We want to decrease the variance, while experiencing some increase in the average length as a benefit of the dropping process. Hence, in Figure 4.7 the minimum increase in average length and maximum decrease in variance values lead us to our objective.

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	100	O	01101
I	001	G	01111
A	011	B	11111
E	0000	C	011101
N	1010	0	01111101
R	0110	1	100111101
U	0101	2	110111101
L	0111	5	011111101
S	01000	3	100011101
K	11000	8	000011101
D	00010	4	001011101
T	10010	9	101011101
M	01110	6	011111101
Y	11110	7	111111101

E = 0.00
Average = 4.01359
Variance = 0.95975

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	010	O	00101
I	001	G	01111
A	011	B	11111
E	0000	C	010101
N	1100	0	000110101
R	0110	1	010110101
U	1101	2	001110101
L	0111	5	101110101
S	01000	3	111110101
K	11000	8	011110101
D	00100	4	0100110101
T	10100	9	1100110101
M	01110	6	0110110101
Y	11110	7	1110110101

E = 0.0005
Average = 4.01413
Variance = 0.96142

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	010	O	00101
I	001	G	01111
A	011	B	11111
E	0000	C	010101
N	1100	0	100110101
R	0110	1	110110101
U	1101	2	001110101
L	0111	5	101110101
S	01000	3	111110101
K	11000	8	011110101
D	00100	4	0000110101
T	10100	9	1000110101
M	01110	6	0010110101
Y	11110	7	1010110101

E = 0.0010
Average = 4.01413
Variance = 0.96142

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	010	O	10001
I	101	G	01011
A	111	B	11011
E	1000	C	011110
N	1100	0	010111110
R	0110	1	110111110
U	1001	2	001111110
L	0011	5	101111110
S	00000	3	111111110
K	10000	8	011111110
D	00100	4	0000111110
T	10100	9	1000111110
M	01110	6	0100111110
Y	00001	7	1100111110

E = 0.0045
Average = 4.01413
Variance = 0.96142

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	010	O	10001
I	101	G	01011
A	111	B	11011
E	1000	C	001110
N	1100	0	010101110
R	0110	1	110101110
U	1001	2	001101110
L	0011	5	101101110
S	00000	3	111101110
K	10000	8	011101110
D	00100	4	0000101110
T	10100	9	1000101110
M	11110	6	0100101110
Y	00001	7	1100101110

E = 0.055
Average = 4.01413
Variance = 0.96142

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	101	O	00010
I	0000	G	10010
A	1100	B	00110
E	1010	C	101101
N	1110	0	01011000
R	0001	1	11011000
U	1001	2	00111000
L	0011	5	10111000
S	1011	3	11111000
K	0111	8	01111000
D	1111	4	000011000
T	01000	9	100011000
M	00100	6	010011000
Y	10100	7	110011000

E = 0.040
Average = 4.07994
Variance =0.48310

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	0100	O	01100
I	1010	G	11100
A	0110	B	00010
E	1001	C	10010
N	0101	0	00001
R	1101	1	10001
U	0011	2	0001110
L	1011	5	1001110
S	0111	3	0101110
K	1111	8	1101110
D	00000	4	1011110
T	10000	9	0011110
M	01000	6	0111110
Y	11000	7	1111110

E = 0.080
Average = 4.24123
Variance = 0.21604

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	1000	O	01100
I	1010	G	11100
A	0110	B	00010
E	0001	C	10010
N	0101	0	011110
R	1101	1	111110
U	0011	2	001001
L	1011	5	101001
S	0111	3	111001
K	1111	8	011001
D	00000	4	0001110
T	10000	9	1001110
M	00100	6	0101110
Y	10100	7	1101110

E = 0.075
Average = 4.14135
Variance = 0.20952

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	0010	O	01100
I	1010	G	11100
A	0001	B	00100
E	1001	C	10110
N	0011	0	01110
R	1011	1	11110
U	0111	2	000101
L	1111	5	100101
S	00000	3	110101
K	10000	8	010101
D	01000	4	001101
T	11000	9	101101
M	00100	6	011101
Y	10100	7	111101

E = 0.15
Average = 4.31951
Variance = 0.22842

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	0000	O	010010
I	1000	G	110010
A	0100	B	001010
E	1100	C	101010
N	0001	0	011010
R	1001	1	111010
U	0101	2	000110
L	1101	5	100110
S	0011	3	110110
K	1011	8	010110
D	0111	4	001110
T	1111	9	101110
M	000010	6	011110
Y	100010	7	111110

E = 0.25
Average = 4.31693
Variance = 0.53337

TABLE 10

CODE WORDS AFTER DROPPING THE LESS FREQUENT SYMBOLS

Symbol	Code Word	Symbol	Code Word
Space	0001	O	01010
I	1001	G	11010
A	0101	B	00110
E	1101	C	10110
N	00000	0	01110
R	10000	1	11110
U	01000	2	00011
L	11000	5	10011
S	00100	3	11011
K	10100	8	01011
D	01100	4	00111
T	11100	9	10111
M	00010	6	01111
Y	10010	7	11111

E = 0.30
Average = 4.55952
Variance = 0.24645

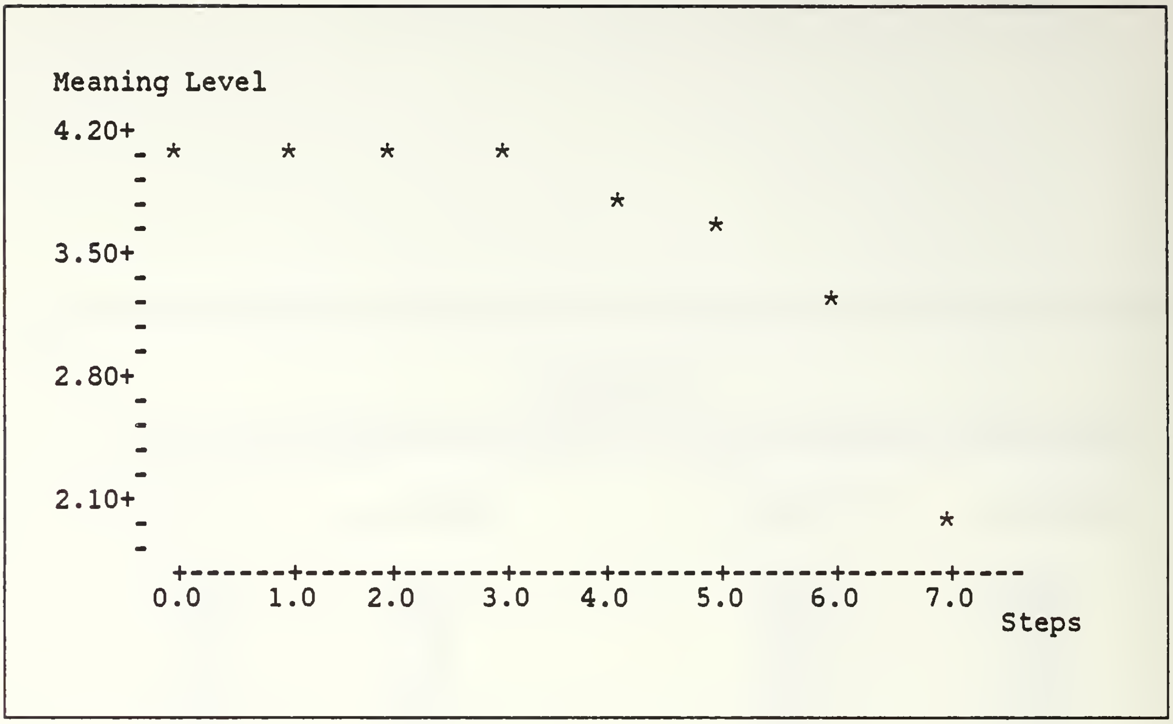

Figure 4.4 Meaning Levels vs. Steps.

E	AVE.	VAR.
0.0000	4.01359	0.95998
0.0005	4.01413	0.96142
0.0010	4.01413	0.96142
0.0045	4.01413	0.96142
0.0055	4.01413	0.96142
0.0400	4.07948	0.48310
0.0800	4.24123	0.21604
0.0750	4.24135	0.20952
0.1500	4.31951	0.22842
0.2500	4.31690	0.53337
0.3000	4.55952	0.24646

Figure 4.5 Average Lengths And Variance At Step 6.

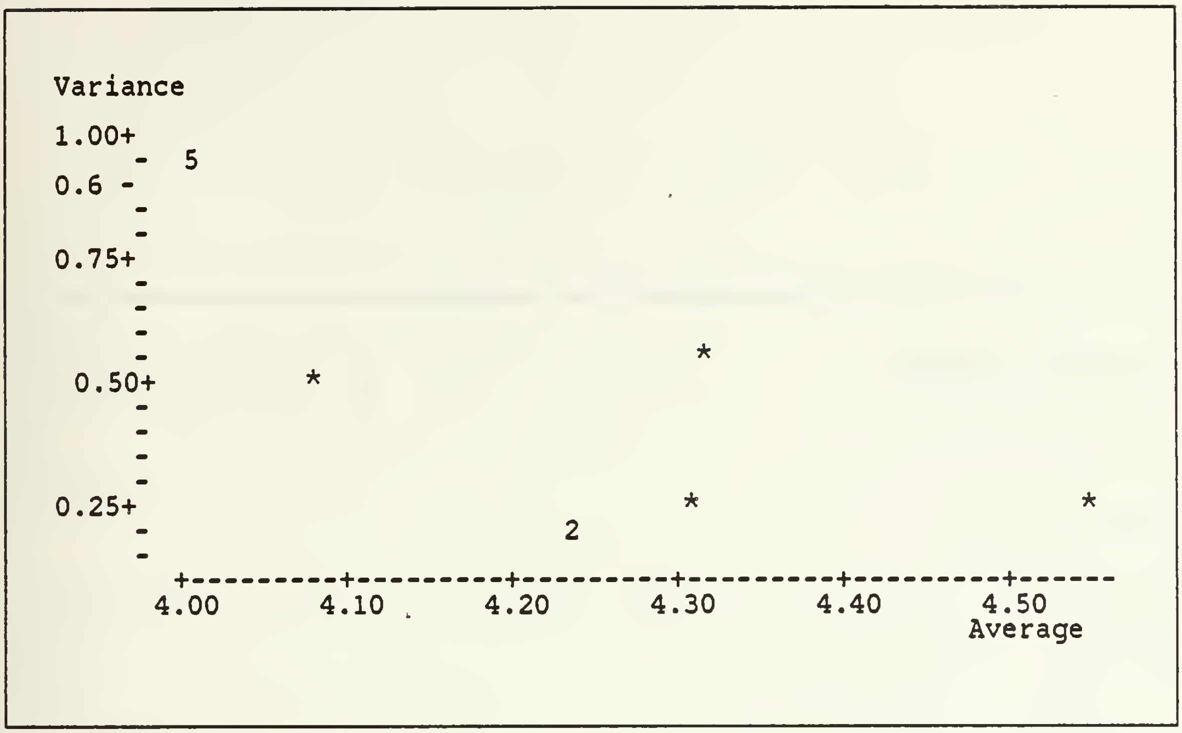

Figure 4.6 Average Length vs. Variance At Step 6.

E	Average Differences	Variance Differences
0.0000	0.294120	0.958220
0.0005	0.294450	0.961470
0.0010	0.297680	0.784060
0.0045	0.297860	0.771469
0.0055	0.311620	0.409499
0.0400	0.304330	0.409000
0.0800	0.258720	0.292300
0.0750	0.352850	0.206540
0.1500	0.414380	0.114560
0.2500	0.580891	-0.365230
0.3000	0.528910	-0.165850

Figure 4.7 Average Length And Variance Differences For C10 And C11.

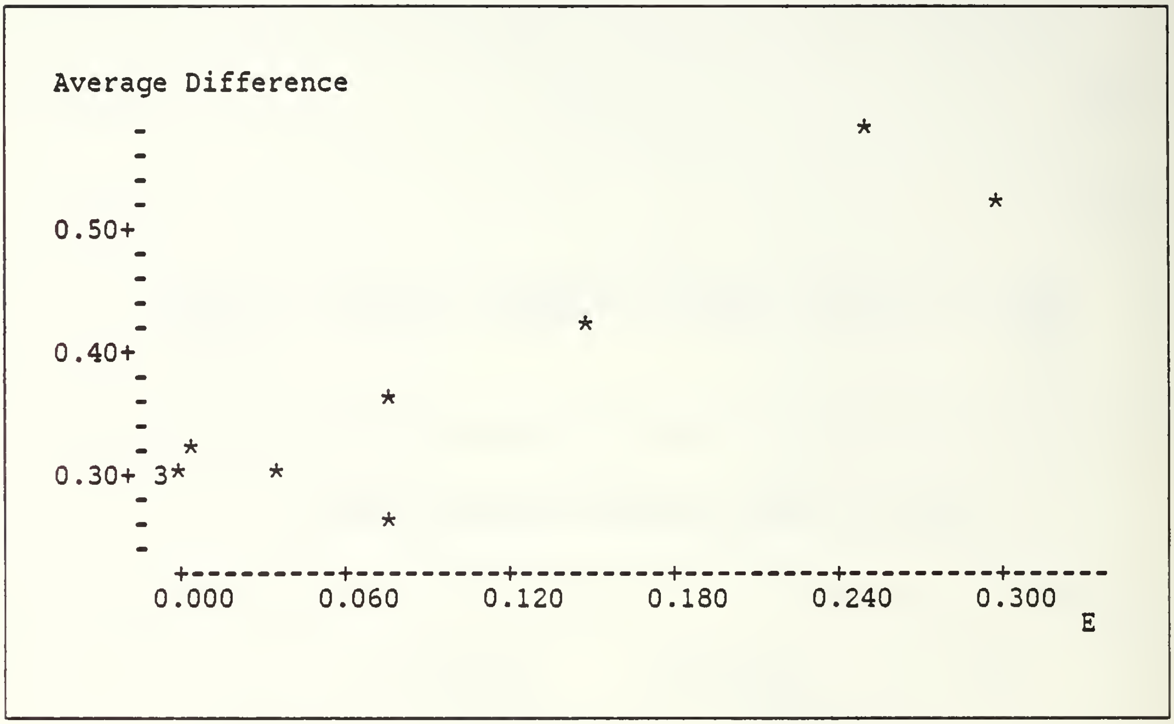

Figure 4.8 Average Length Differences vs. E Values.

```
Variance Difference
     - 2
     -
0.80+ **
     -
     -
     -
     -
0.40+  *     *
     -              *
     -              *
     -
     -                          *
0.00+
     -
     -                                                   *
     -
     -
-0.4+                                           *
       +---------+---------+---------+---------+---------+-----
     0.000     0.060     0.120     0.180     0.240     0.300
                                                               E
```

Figure 4.9 Variance Differences vs. E Value.

V. REDUCTION IN BANDWIDTH

A. QUEUEING THEORY IN A COMMUNICATIONS SYSTEM

In communications systems, as with other systems, the managerial view is to use available sources effectively. The limited availability of the frequencies, according to which transmission bandwidths are determined, makes the managerial job harder. The main idea is to transmit in a way that uses the minimum required bandwidth. This adjustment is done by employing the output rates which satisfy the objective.

Our communications system model has its own characteristics. The input rate of the system is the number of digits which comes to the transmitter in a unit time. This number of digits is the combination of the 0's and 1's. The frequency distribution of arrivals is similar to the theoretical Poisson distribution. As described in [Ref. 8,] the Poisson distribution occurs when arrivals are random in a given period of time (unit time). This means that although we know the mean arrival rate for unit time, the exact arrival can not be predicted at any given moment. Input rate is represented by lambda (λ).

The output rate is the number of digits which are transmitted in a given period of time (unit time). The inverse of the output rate is described as the output time and has the negative exponential distribution. The output rate is represented with mu (μ).

In our model, the input rate is diretly affected by the symbols which comprise the messages and are intended to be sent through our communications system. Since each symbol has its own digital representation after the encoding process (Modified Huffman coding), the number of digits which enter the system varies. In other words, the number of digits has the lowest and the peak values.

As a system manager, we can adjust our output rate in two ways. The first way is to have an output rate as high as the peak value of the input. But, since the peak value doesn't always occur, the resources will be wasted. The second option is to set the output rate close to the mean input rate and then store the excess digits in a buffer. Having a buffer allows us to reduce the output rate. The reduction in the output rate also means having a gain in the transmission bandwidth. These two together allow management to achieve effective usage of the resources.

At this point, the size of the buffer should be considered. The size of the buffer directly affects the efficiency of the system. Gains in bandwidth should result in buffer size increase, i.e. having a very large buffer.

Every digit is transmitted according to an arrival sequence. The first digit arriving is always the first digit transmmitted, and so on. This technique is called first - in - first - out (FIFO.) This guarantees that the sequence, the meaning of the symbols and the meaning of the messages are not destroyed.

Another characteristic of the system is that it has just one transmitter. Every digit which comes into the system is transmitted through one channel. If the output rate is smaller than the input rate, the buffer size will increase without bound. So, the input rate should be less than or equal to the output rate ($\lambda <= \mu$.)

In summary, our communications system model can be represented as M / M / 1. This representation means, single transmitter (1), Poisson arrivals (M), and exponential output (M.)

B. SIMULATION OF THE COMMUNICATIONS SYSTEM

1. Simulation Without Dropping Process

The communications system has been simulated with the computer programs written in Pascal. The first program, given in Appendix F, is the simulation of the communications system which encodes the Turkish alphabet symbols without dropping any of them. This program was run nine times with nine different codes. Eight of these codes were chosen among the eleven codes which were given in Table 6. For the ninth run, the "block code" was used. They are given in Figure 5.1.

The output of the program is the maximum buffer size which is required for transmitting the first 200 characters of the article given in Appendix A. The output rates are chosen arbitraily and they are 4.01359, 4.2, 4.4, 4.6, 4.8, 5.1, 5.5, and 6.0. The maximum output rate, 6.0, is the rate that is required to transmit the block codes without any buffer.

The output rates which are lower than 6.0 represent the gain in the bandwidth. The required maximum buffer sizes for each code are given in Figure 2.1 and each output rate are given in Table 11. The gains in bandwidth at each output level are given in Figure 5.2.

When Table 11 is examined the following results can be concluded:

(1) As the gain in bandwidth increases (lower output rates), the maximum required buffer size also increases.

(2) Modification of Huffman coding, (increase in average code length and decrease in variance) gives us a lower buffer size. (This conclusion is true when $\lambda < \mu$.)

E	Ave. Len. (Input R.)	Code Name
0.0	4.30771	CodeA1
0.005	4.30858	CodeB1
0.040	4.38381	CodeC1
0.080	4.49995	CodeD1
0.075	4.59420	CodeE1
0.150	4.73389	CodeF1
0.250	4.89779	CodeG1
0.300	5.08843	CodeH1
	6.0	CodeI1

Figure 5.1 Code Lengths And Code Names.

Output Rate	Bandwidth
6.0	0.0
5.5	8.30
5.1	15.00
4.8	20.00
4.6	23.33
4.4	26.66
4.2	30.00
4.01359	33.11

Figure 5.2 Percentage Gain In Bandwidth For Each Output Rate.

2. Simulation With Dropping Process

The second experiment is the simulation of the communications system which encodes the Turkish alphabet after dropping the same source symbols which have probabilities less than the optimal limit probability (P = 0.015.) This program was also run nine times, applying nine different codes. Eight of these codes are chosen among the eleven codes which were given in Table 11. Again, block coding is used as

TABLE 11

MAXIMUM REQUIRED BUFFER SIZES WITHOUT DROPPING

Input Rate	Output Rate								Code
	4.01359	4.2	4.4	4.6	4.8	5.1	5.5	6.0	
4.30771	92	64	39	30	23	21	18	14	A1
4.30858	92	64	39	30	23	21	18	14	B1
4.38831	86	52	24	18	16	14	11	8	C1
4.44995	107	70	34	17	13	11	8	5	D1
4.59420	123	86	47	17	8	6	4	1	E1
4.73389	144	107	67	29	9	6	4	1	F1
4.89779	178	141	103	61	22	5	3	1	G1
5.08843	222	184	144	104	64	9	3	0	H1
6.0	398	360	320	280	240	180	90	0	I1

the ninth code. These nine codes' average lengths and the code names are given in Figure 5.3. The computer program is given in Appendix G.

The output of the system is the same as the first program. The maximum required buffer size for the first 200 characters of the article is given in Appendix A. The same output rates which are given in Figure 5.2 were used.

The maximum buffer size requirements for each code, at eight different output rates are given in Table 12. After examining Table 12, we can reach the same conclusion as we did in the first section, that is: higher average length results in smaller buffer size.

We must compare the two tables (Table 11 and Table 12) in order to find out the effect of the dropping process in bandwidth and buffer size. In order to compare before and after dropping encoding processes easily, the output rates are chosen to be exactly the same in both models. The minimum value of the output rate is the minimum average length value, reached after the dropping process. It is named "code A2." It is actually the Huffman coding process result after dropping, because the applied E value is 0.0. This output rate gives a 33.11% gain in bandwidth when compared with the output rate (6.0).

It can easily be seen that the buffer sizes which are the results of the modified Huffman coding process after dropping the less frequent symbols are much smaller

than the before (without) dropping process. For example, let's compare buffer sizes which are the results of codes D1 and D2. Both D1 and D2 used the same modification parameter E = 0.08. Code D1 was done without dropping and code D2 was done after dropping. Since the buffer sizes are reached by employing the same output rates, both D1 and D2 have the same percentage bandwidth gain for each output rate. This results in a reduction/gain in buffer size in addition to a gain in bandwidth. The buffer sizes and the percentage gains of Code D2 are given in Figure 5.4.

By looking at the percentage buffer size gains of Code D2 at each output level, we can easily see the positive effects of the dropping process on the buffer size. Between Code D1 and Code D2, the mean percentage gain is 75.35%.

Hence, it can easily be concluded that besides a maximum 33.11% reduction in bandwidth, the dropping process can give us an average of 75.35% reduction in buffer size. The change in buffer size for our sample codes D1 and D2, while increasing the output rate (decreasing the bandwidth gain) can be seen in the graph given in Figure 5.5. The vertical axis is the maximum required buffer size. The horizontal axis is the output rate, with the bandwidth gain given in parentheses. The area between curves D1 and D2 gives the buffer size gain of Code D2.

TABLE 12

MAXIMUM REQUIRED BUFFER SIZES WITH DROPPING

Input Rate	Output Rate 4.01359	4.2	4.4	4.6	4.8	5.1	5.5	6.0	Code
4.01359	27	20	18	17	16	14	11	9	A2
4.01413	31	24	22	21	20	18	15	12	B2
4.07948	28	19	17	16	15	13	10	8	C2
4.24123	50	16	6	4	3	2	2	1	D2
4.27135	53	19	9	7	6	4	2	0	E2
4.31951	64	29	5	3	2	1	1	0	F2
4.31690	78	44	14	7	6	4	2	0	G2
4.55952	105	70	33	4	2	0	0	0	H2
6.00	372	337	300	262	225	169	94	0	I2

E	Ave. Len. (Input R.)	Code Name
0.0	4.01359	CodeA2
0.005	4.01413	CodeB2
0.040	4.07948	CodeC2
0.080	4.24130	CodeD2
0.075	4.27135	CodeE2
0.150	4.31951	CodeF2
0.250	4.31690	CodeG2
0.300	4.55952	CodeH2
	6.0	CodeI2

Figure 5.3 Code Lengths And The Code Names.

Output Rate	D1 Buffer	D2 Buffer	Gain (%)
4.01359	107	50	53.27
4.2	70	16	77.14
4.4	34	6	82.35
4.6	17	4	76.47
4.8	13	3	76.92
5.1	11	2	81.81
5.5	8	2	75.00
6.0	5	1	80.00

Figure 5.4 Buffer Size Gain Of Code D2.

```
Maximum Buffer Size

     -
     -
 105+     D1
     -
     -
     -
     -
  70+             D1
     -
     -
     -     D2
     -
  35+                     D1
     -
     -
     -            D2              D1      D1          D1
     -                    D2      D2                              D1                  D1
   0+                                     D2          D2          D2                  D2
      ----+-------+-------+-------+-------+-----------+---------------+------------------+--
   4.01359      4.2     4.4     4.6     4.8         5.1             5.5                6.0

                                                                          Output Rate

(%)(33.11)  (30)  (26.6)(23.3)(20.0)  (15.0)     (8.3)          (0.0)
```

Figure 5.5 Max. Buffer Size of The Codes D1 And D2.

VI. TWO ALTERNATIVE APPROACHES

A. DROPPING MORE FREQUENT SYMBOLS

During my experiments, another approach to the dropping process appeared logical, namely to drop the more frequent symbols instead of the less frequent symbols. The theoretical explanation and an example of this experiment are given below.

Since the main idea is to decrease the variance of the code words, the smaller variance can be reached mathematically, not just by avoiding longer length code words, as explained in section 2.3, but also by avoiding larger symbol probabilities. If we recall the variance formula:

$$V = \sum P_i \, (X_i - L_{ave})^2$$

the smaller P_i's are, the smaller the variance is.

When we apply this explanation in our coding process, the purpose is to leave the higher probability symbols out of the calculation. In other words, drop them before encoding.

The average length and variance results after dropping the more frequent symbols for the same example given in Figure 3.3 is shown below.

$$L = 0.7$$

$$V = 0.716$$

When we encode the same sample source alphabet, after dropping the more frequent symbols, by using Modified Huffman coding for $E = 0.2$, the results are changed.

$$L = 0.8$$

$$V = 0.576$$

Average lengths and variances for four different encoding processes are summarized in Table 13. When we examine columns two and four, it can be said that dropping the more frequent symbol has a greater effect on average length than it has on variance. Although encoding after dropping the more frequent symbols results in a high decrease in average length, it gives an increase in variance if we compare with Modified Huffman coding variance (0.24.) But, since the decrease in average length is very high, we can still apply this idea in our sample messages.

Symbol	P	Symbol	P
S1	0.4	S3	0.2
S2	0.2	S4	0.1
S3	0.2	S5	0.05
S4	0.1	S6	0.05
S5	0.05		
S6	0.05		

Figure 6.1 Dropping Process.

Symbol	P	P	P
S3	0.2(0)	0.2(0)	0.2(0)
S4	0.1(10)	0.1(10)#	*0.2(1)
S5	0.05(110)#	*0.1(11)#	
S6	0.05(111)#		

Figure 6.2 Huffman Coding After Dropping.

Symbol	Code Word
S3	1
S4	10
S5	110
S6	111

Figure 6.3 Final Code Words.

Symbol	P	P	P
S3	0.2(00)	*0.3(1)	*0.5(0)
S4	0.1(01)	0.2(00)#	0.3(1)
S5	0.05(10)#	0.1(01)#	
S6	0.05(11)#		

Figure 6.4 Modified Huffman Coding After Dropping for E=0.2.

Symbol	Code Word
S3	00
S4	01
S5	10
S6	11

Figure 6.5 Final Code Words For E=0.2.

TABLE 13

RESULTS OF THE FOUR DIFFERENT CODING TECHNIQUES

	Huffman	Modified H.	Huffman C. After Drop.	Modified H.C. After Drop.
Aver.	2.3	2.4	0.7	0.8
Var.	1.81	0.24	0.716	0.576

B. DROPPING MORE FREQUENT SYMBOLS FROM TURKISH ALPHABET

The idea mentioned in the prior section is applied to the sample articles which were given in Appendix D. These four sample articles are rewritten several times. At the first step, the second most frequent symbol, I, is dropped. Then, at each step one more symbol, in decreasing frequency order, is dropped. At the seventh step, a total of seven symbols (I, A, E, N, R, U, L) are dropped.

After this last step, some other symbol combinations, which were chosen among the more frequent ones, are dropped. The step numbers and the dropped symbols on each step are given in Table 14.

Examination of the rewritten articles showed that the meaning of the articles were dramatically destroyed at every step after the first one.

According to the author's observation, dropping more than one of the vowels affected the meaning level significantly. So, for the sake of experiment, these articles were rewritten by dropping one vowel symbol and one or more consonant symbols. The symbol which was used at the first step, I, is chosen as the vowel symbol.

First, the (I, N) combination, followed by the (I, N, R) combinations were dropped. The desired meaning level was reached with the first set.

After choosing the symbols which can be dropped without destroying the meaning of the articles, the Modified Huffman coding process is applied. The resulting average lengths and variances for each experimental E value are given in Figure 6.6. The articles after dropping I and N are given in Appendix H.

The Huffman encoding without modification (E = 0.0) and without dropping any symbols, gives an average length of 4.30771 and a variance of 1.9182 (Figure 3.6.) When we compare these values with Figure 6.6 values, it can easily be seen that dropping more frequent symbols (I and N) gives us a reduction in both average code length and variance. Hence, dropping the more frequent symbols is also a solution to reducing bandwidth and buffer size.

TABLE 14

DROPPED SYMBOLS AT EACH STEP

Step No.	Dropped Symbols
1	i
2	i a
3	i a e
4	i a e n
5	i a e n r
6	i a e n r u
7	i a e n r u l
8	i n r u l
9	i n u l
10	a r l u
11	a e r s
12	a l s k

E	Average L.	Variance
0.0	4.19743	1.77221
0.0005	4.19881	1.78210
0.0010	4.20002	1.79594
0.0450	4.21403	1.55652
0.0550	4.22123	1.56867
0,0400	4.34790	1.79754
0.0800	4.40201	0.60670
0.0750	4.39815	0.53646
0.1500	4.65842	0.31867
0.2500	4.66691	0.31431
0.3000	5.04798	0.04568

Figure 6.6 Average Lengths And Variances After Dropping I And N.

C. DROPPING "MORE AND LESS" FREQUENT SYMBOLS TOGETHER

The positive effect of dropping less frequent symbols on the bandwidth and the buffer size is shown in Chapters 3 and 5. The same positive effect of dropping more frequent symbols is shown in the previous section. Since both approaches yield the desired result separately, could they give the same desired result when applied together? In other words, if we drop some more frequent symbols and some less frequent symbols, what would the result be?

To answer this question, at each step in addition to the more frequent symbols examined in the last section (I and N), some less frequent symbols were dropped and the same articles were rewritten. Figure 6.7 gives the step numbers and the corresponding symbols.

The meaning level of the articles is destroyed after the fourth step. The more and less frequent symbols combination which can be dropped without destroying the meanine is (I, N, Q, X, ?, -, :, W, J, ;, (,), ", ', ., ,). The modified Huffman encoding process is applied after dropping this more and less frequent symbols combination. The resulting average lengths and variances for each experimental E value are given in Figure 6.8. The rewritten forms of the articles are given in Appendix I.

This process not only gave smaller average lengths and variances than Huffman coding, but also gave smaller values than dropping more frequent symbols (Figure 6.6.) Hence the gain (reduction in the bandwidth and the buffer size) is more than the gain of the Huffman coding, modified Huffman coding and modified Huffman coding after dropping more frequent symbols.

It should be mentioned here that the optimal symbol combination which can be dropped without destroying the meaning level of any message is the subject of a more detailed research. It, of course, changes from alphabet to alphabet and also changes from field to field. For example, this combination might be different in the military than in chemistry or some other field. Also, the frequent usage of the same words and phrases in one field might let users of this field drop more symbols for communicating within this field.

Step No.	Dropped Symbols
1	i n q ? - :
2	Step 1 symbols and w j
3	Step 2 symbols and ; ( )
4	Step 3 symbols and " , .
5	Step 4 symbols and f h

Figure 6.7 Dropped Symbols At Each Step.

E	Average L.	Variance
0.0	4.04584	1.41666
0.0005	4.04618	1.28315
0.0010	4.04618	1.28315
0.0450	4.05852	1.22259
0.0550	4.05852	1.22259
0,0400	4.26919	0.58460
0.0800	4.29284	0.28517
0.0750	4.31350	0.30860
0.1500	4.61841	0.24088
0.2500	4.48118	0.26255
0.3000	4.83324	0.13819

Figure 6.8 Average Lengths And Variances After Dropping.

VII. EVALUATION OF THE RESULTS AND CONCLUSIONS

A. EVALUATION OF THE RESULTS

As stated earlier, the main objective of this research is to reduce the bandwidth, in other words the transmission bit rate, and the buffer size. The technique used is to drop the less frequent source symbols before encoding and then to encode the message by using Modified Huffman Coding.

The results which were reached in Chapter 4 gave a 33.11% reduction in bandwidth and a 75.35% reduction in buffer size, for the first 200 symbols of the given message in Appendix A. The gain calculation was done by comparing the results of encoding after dropping less frequent symbols with the results of the block coding process.

During experiments for finding the optimal modification parameter (E) after the dropping process, it was concluded that a change in the number of symbols in the source alphabet affected the optimal modification parameter. In other words, a modification parameter which is optimal for a given source alphabet is not necessarily optimal for the same alphabet after the dropping process. This conclusion leads us to calculate the optimal modification parameter individually for each separate source alphabet with a given number of symbols.

In Chapter 6, two additional approaches were briefly examined. The first is to encode the message after dropping the more frequent symbols by using Modified Huffman Coding. The second one is to encode after dropping more and less frequent symbols combination. In order to show the effect of these last two approaches on the average code lengths and variances, let's choose a modification parameter and compare the resulting values of these parameters with the Huffman coding results without modification and dropping.

In Chapter 3, for E = 0.04 the average code length and the variance without dropping any symbols were calculated to be 4.38381 and 0.89210 (Figure 3.6.) After dropping the less frequent symbols, for the same parameter the results are 4.07948 and 0.048310 (Figure 4.5.) Finally, in Chapter 6, after dropping the more frequent symbols the results are 4.3479 and 1.79454 (Figure 6.6.) And, from Figure 6.8, the average code length and variance after dropping the more and less frequent symbol

combination are 4.26919 and 0.58460. If we recall the Huffman coding results, without modification and without dropping, from Chapter 3, Figure 3.6, the average length is 4.30731 and variance is 1.91820. Figure 7.1 shows these values together. We rearrance Figure 7.1 in decreasing variance order, as shown in Figure 7.2. Figure 7.3 shows the same code results in decreasing average length order.

As explained earlier, longer average length means a larger bandwidth is required. When Figure 7.3 is examined, it can be seen that the dropping processes decrease the average length. Although the fourth code has a larger average length than Huffman coding, it is smaller than modified Huffman coding average length. This is the effect of the modification process on the average length. The modification process, without dropping any symbol, while decreasing the variance, increases the average length.

At this point, when Figure 7.2 is examined, it is seen that this negative effect of the modification process (increase in average length, while having a decrease in variance) is eliminated by using codes five and three. These two, encoding after dropping more and less frequent symbols and after dropping less frequent symbols by using Modified Huffman Coding, not only decrease the variance but also decrease the average length as well.

Code No.	Code Name	Average L.	Variance
1	Huffman	4.30771	1.19820
2	Modified H.C.	4.38381	0.89210
3	Drop. Less Fr. M.H.C.	4.07948	0.48310
4	Drop. More Fr. M.H.C.	4.34790	1.79454
5	Drop.'More&Less'M.H.C.	4.29919	0.58460

Figure 7.1 Results Of Five Different Coding For E=0.04.

Code No.	Code Name	Average L.	Variance
1	Huffman C.	4.30771	1.91820
4	Drop.More Fr.M.H.C.	4.34790	1.79154
2	Modified H.C.	4.38381	0.89210
5	Drop.'More&Less'M.H.C.	4.29919	0.58460
3	Drop. Less Fr. M.H.C.	4.07948	0.48310

Figure 7.2 Results In Decreasing Variance Order.

Code No.	Code Name	Average L.	Variance
2	Modified H.C.	4.38381	0.89210
4	Drop.More Fr.M.H.C.	4.34790	1.79154
1	Huffman C.	4.30771	1.91820
5	Drop.'More&Less'M.H.C.	4.29919	0.58460
3	Drop. Less Fr. M.H.C.	4.07948	0.48310

Figure 7.3 Results In Decreasing Average Length Order.

B. CONCLUSION

We have shown that employing Modified Huffman coding after dropping the less frequent or more and less frequent source symbols combination results in a decrease in variance as well as in average code length.

A decrease in average length reduces the number of digits transmitted in a unit time in a communications system. Hence, this communications system can either handle the same amount of traffic with less transmission bandwidth, and share the excess capacity with others, or with the same available bandwith transmit a greater traffic load. In both cases, the required buffer size, due to a dramatic reduction in the variance, will be very small. This reduction in buffer size results in a cost savings as well as reduces the need for complex network flow control algorithms.

In addition to these benefits of the dropping process, the Modified Huffman coding technique can be used for encryption (since each E value results in a unique set of code words.) The modification parameter E can be considered as an encryption key and distributed for each encryption period to the stations for decryption. This presents a subject for future research.

APPENDIX A
THE TURKISH MAGAZINE ARTICLES

1. "STRANGE SHAPES OF MODERN SHIPS"

The first article titled "Strange Shapes of Modern Ships" is given below.

Bir derginin ressami, en guclu vinclerin yapamadigi isi basararak, 50.000 tonluk bir "olyanus devi"ni Sudan cikardi ve boylece, geminin burnundaki yumrubas "balb" ortaya cikmis oldu. Geminin kic tarafinda da bazi yenilikler goze carpiyordu. Bunlarin sirri acaba ne olabilirdi? Otomobil yapimcilarinin yeni gelistirdikleri modelleri denedikleri "ruzgar tunelleri"nin bir benzeri deniz tekneleri uzerinde calisan meslektaslari icin de gecerli oluyor. Onlarin da yeni tekne modelleri denedikleri "test havuzlari" var. Yeni gemiler, ancak, bu havuzlarda yapilan deneylerin olumlu sonuclar vermesinden sonra, insa edilmek uzere kizaga konuyor. Bu arada, gemi muhendislerinin isleri, kara araclari uzerinde ugras veren meslektaslarinin islerinden biraz daha guc. Bu gucluk, daha model asamasinda baslar. Deneyleri yapilan gemi modelleri, yeterince buyuk oldugu zaman, deneylerden alinan olcum sonuclari, istenileniverebilmektedir. Guclugu yaratan ikinci etken de, dunyamizin "su" ve "hava" olarak bilinen iki elamamindan kaynaklanmaktadir. Bir kara tasitinda, laroseri sadece ruzgara karsi koymak zorunda olmasina karsin, bir teknenin hem dalgaya ve hem de, ruzgara karsi koymasi gerekir. Eski tarihlerde insa edilmis gemilerde, burunlar keskinlestirilir ve boylece suyun daha az bir direnimle yarilmasi saglanirdi. Ancak, bu is, aslinda hic de gorundugu kadar basit degildir. Gemi hesaplari, sualtindan ateslenen bir roketin hesaplarindan daha karmasik ve gictur. Biraz once belirtigimiz gibi bir gemi, su ve hava ortaninda seyreder. Bu nedenle de, ozellikle havanin ve suyun birlestigi nokta, muhendisler icin bir "bilmece"dir. Beney havuzlarindan alinan sonuclar okyanuslar icin de gecerli oldugundan; bubenzer iliskilerden yararlanan gemi m uhendisleri, deneylerini deney havuzlarinda yapmaktadirlar. Gemiye hareket veren pervane, tekneyi ileriye iterken, geminin burnunda bir dalga olusur. Bu dalga, burunda, yanlarda, dipte ve kicta gemiyi yalayarak gecer. Ancak, anilan dalga alisilagelen tipte bir dalga olmayip, saga-sola karisik hareketler yapan sular halindedir. Gemi burnunda olusan ve tekne tarafindan iletilen bu su kitleleri, gemi burnunun genisligi oraninda artan bir yigilma yaparak, istenilmeyen bir direnc olusturur (sekil 1). Istenilmeyen bu direncin etkisini azaltabilmek icin, geminin burnunda yumrubas denilen ve mahmuzu

andiran bir cikinti yapilir. Yumrubasin etkisi soyle aciklanabilir: yumrubasli bir tekne, onunde iki dalga tepesi olusturur. Bunlardan, teknenin olusturdugu dalga tepesi, yumrubasin olusturdugu dalganin cukurunun doldurarak, gemi burnundaki yigilmayi onler (sekil 2). Donuc olarak da, istenilmeyen dalga yok edilir. Yumrubas adi verilen bu yeni burun tipi, Amerikali gemi adami David Taylor'un bulusudur. Yuzyilimizin baslarinda Taylor, yumrubasli gemilerin, digerlerine kiyasla daha kucuk dalgalar olusturdugunu tespit etmis ve bunun teorisi saha sonra gelistirilmistir. Ancak, tum olasiliklari aydinliga kavusturacak kesin formuller gunumuzde dahi tam olarak saptanmis degildir. Yumrubas teorisinin gelismesini asagidaki maddelerle acikliyabiliriz: (1) seyir halindeki bir gemi, onunde buyuk bir dalga tepesi olusturarak ilerler. (2) su yuzeuinin hemen altinda hareket ettirilen bir kure, arkasinda bir dalga cukuru olusturur. (3) gemi modelinin burnuna bir kure yerlestirilerek, kurenin olusturdugu dalga cukuru ile gemi modelinin olusturdugu dalgayi cakistiracak bir deney uygulamasi gerceklestirilir. (4) deneyde, dalga cukurunun dalga tepesini yuttugu gorulur. (5) dalga tepesi yutuldugundan; istenilmeyen direnc etkisini kaybeder. Sonuc olarak, gemi modeli daha buyuk bir hiz kazanir veya hareketi icin gerekli olan guc azalir. Alinan bu sonuc, geminin tukettigi yakitta hic de azimsanmayacak bir tasarruf saglandigini ortaya koyar. Armatorlerin yumrubasli gemi siparislerine agirlik vermelerinden sonra, muhendislerin isleri daha da guclesmistir. Ilk zamanlarda yumrubaslar, yolcu ve savas gemilerinde uygulaniyordu. Bununda nedeni, anilan gemilerin seferlerini genellikle sabit bir su kesiminde yapmalari idi. Oysa, armatorun siparise bagladigi yuk gemilerinde su kesimi (draft), gemilerin yuklu veya bos olmalarina gore, degisebildigi icin, gemi burnunda yer alal yumrubas, etkinlik pozisyonunu koruyamamaktadir. Gemi, yukunu alarak sefere ciktiginda; yumrubas, sualtinda, kalarak, etkinligini surdurmekte ise de, yukun bosaltilmasindan sonra, su yuzeyine cikmakta ve sonuc olarak, etkinligini kaybetmektedir. Bu durum, yumrubasin gemi burnunda nerede yer almasi gerektigi sorununu ortaya cikarmistir. Daha sonra, yumrubas, gemi burnunun biraz daha asagisina alinarak, suyun altinda birakilmis ve istenilen sonuca kismen de olsa ulasilmistir. Yumrubasi sadece sualtinda birakmakla sorunlara cozum getirilememektedir. Cunku, her tekne kendine ozgu bir dalga sekli olusturmakta be bu nedenle de, yumrubasin, kullanacagi tekne ile uyum saglayacak ozelliklere sahip olması gerekmektedir. Gemi muhendislerinin goguslemek zorunda olduklari bu guclukler, yenbi arastirma alanlarinin dogmasina yol acmis ve bu kez de, arastirmalar geminin kic tarafinda yogunlasmistir. Yaklasik 20 yil kadar once, Hamburglu gemi muhendisi ernst

nonnecke, yeni bir kic formu gelistirmis ise de, onun bu bulusu ancak son yillarda deger kazanmaga ve dikkat cekmege baslamistir. Nitekim, nonnecke'nin bulusu, bir kore tersanesinde 2 konteyner gemisinde uygulamaya konulmustur. Teorik calismalar Hamburg'da baslamis ve bunu izleyen deneylerde, insa edilecek geminin bir modeli, boyu 300 M. Ve derinligi 18 M. olan bir deney havuzuna cekilerek, nonnecke'nin gelistirdigi kic formunun ustunlugu kabul edilmistir. Bu tip asimetrik kic formu: sancak tarafi cukur ve iskele tarafi disa dogru bombelidir. Bu formun ozelligi, suyun akisini duzelterek, dogrudan pervaneye vermesidir. Nonnecke tipi kic teorisi su sekilde aciklanabilir: sivi icinde hareket eden bir govde, suyu bas taraftan yarar. Yarilan su, govdenin kic tarafinda yine birlesmek egilimi gosterirken, bu kez de geminin pervanesi ile karsilar. Geminin hareket yonune gore, saga dogru donen pervane, suyu teknenin sancak (sag) tarafindan asagiya iter, buna karsin, iskele tarafindan (sol), yukariya dogru itilerek, teknenin kic tarafinda birleseme egilimi gosteren su, birlesmeden pervanenin akimina kapilir. Cekilen sualti fotograflari ile tespit edilen bu olay, suyun gemide iskele tarafindan gerektirdigi itici gucu olusturmadan, yukariya dogru itildigi gercegini ortaya koymustur. Bu olay uzerinde duran nonnecke, iskele tarafindan pervaneye yonelen su akisini duzenleyebilmek icin gemide sancak be iskele taraflarinin pervaneye yakin olan kisimlarinda, tasarladigi form degisikliklerini gerceklestirmistir. Buna gore, geminin sancak tarafi cukurlastirilmis; iskele tarafinda ise, cukurlugun yerini yumusak bir bombe almistir (sekil 5). Sonuc olarak,suyun dagilmaksizin ve turbulansa ugramaksizin, pervaneye akabilmesi saglanmistir (sekil 3 ve 5) eski ve yeni tip iki geminin en kesit egrilerini vermektedir. Eski tip bir gemide en kesit egrileri simetrik bir bicim gostermekte ve geminin ortasinda duz bir cizgi boyunca birlesmektedir (sekil 3). Diger tip kic formunda ise, anilan egriler a simetrik olarak gelmekte ve geminin ortasinda "S" sekilindeki bi cizgi uzerinde toplanmaktadir (sekil 5). Sekil 4 ve 6'da, eski ve yeni tip kic formlarinin birer profili ile pervaneye dogru yonelen suyun akisi gorulmektedir. Eski tip kic formunda (sekil 4); pervaneye dogru akis yapan su, pervane ile karsilastiginda turbulansa ugramakta ve dolayli olarak da, gemi dieselinin pervaneye aktardigi gutce kayba yol acmaktadir. Nonnecke tipi kic formunda ise, pervaneye yonelen suyun akisi duzenlenmis (sekil 6) ve duzenlenen su, turbulansa ugramadan, pervane tarafindan itilerek, pervanenin verimi artirilmis ve geminin daha az bir gucle daha buyuk bir hiz kazanmasi saglanmistir. "Thea S" adli 124 metrelik gemide yapilan deneyler, bu yeni kic formunun gunde 2.000 litrelik bir yakit tasarrufu sagladigini ortaya koymustur. Eski tip gemi formlarinin gecerli oldugu gunlere kiyasla, yakit

fiatlarinin bugun 10 kat arttigi goz onunde tutulursa, gemilere saglanlan yakit tasarrufunun ne kadar onemli oldugu ve modern gemilerinin nicin boyle garip bicimlerde insa edildigi sorusu kendiliginden aydinliga kavusabilir.

2. "STORY OF THE SPACE SHUTTLE

The second magazine article is titled "Story of the Space Shuttle" and is given below.

1970'lere dek dayanan uzay mekigi projesinin temel amaci, uzaya daha ucuz ve dolayisiyla daha sik gitmektir. Mekikten once uzaya atilan insanli ve insansiz uydular, sonda ve roketler sadece bir kez kullanilabiliyordu ve bu nedenle maliyetleri yuksek oluyordu. Uzay mekigi projesi ile insanoglu, ayni uzay aracini surekli kullanma olanigina kavustu. Bu projenin en belirgin ozelligi ucak teknolojisi ile uzay teknolojisini bir araya getirmesidir. Sistem genelde uc ana bolumden olusmaktadir: (1) yorunge araci da denen uzay gemisinin kendisi; (2) buyuk dis yakit tanki; (3) dis yakit tankinin her iki tarafinda bulunan kati yakitli roketler. Sistemi firlatma aninda, geminin arkasinda bulunan ana motorlar ve iki firlatici roket ateslenir. Bu islemin sonunda, otuz milyon newton'luk cok buyuk bir firlatma kuvveti, sistemi havalandirir. Havalandiktan bir dakika sonra sistemin surati, ses suratini asar. Bu sirada geminin icinde olsaniz ve kendinizi tartsanis, yeryuzunde 60 kilo gelen vucudunuzun, iki dakika incinde sismanlamis olmamasina karsin, 180 kilo geldigini gorursunuz. Bu ilginc durum, aracin ivmesinin, cekim ivmesinden uc kat fazla olmasindan kaynaklanmaktadir. Havalandiktan sonra kati yakitli roketlerin yakitlari biter ve dis yakit tankindan ayrilirlar. Bu anda gemi, 50 km. Yukseklikte ve hizi Saatte 5.000 km'ye ulasmistir. Ayrilan roketler, ilk hizlarindan dolayi derhal asagiya dusmezler. 50 km'de ayrilan bu roketler, 67 km'ye dek cikar ve sonra dusmeye baslar. Duserken, yuzeyden yaklasik 3 km. Yukseklikten, uc evreli parasut sistemi calisir ve dususun hizini azaltir. Denize dusen roketler, su yuzeyine degdikleri anda parasutlerden ayrilir ve alt tarafta bulunan ozel bolmeler siserek, roketlerin batmamalari saglanir. Daha sonra bunlar denizden toplanir. Gerekli onarim ve bakim yapilarak, bir sonraki ucus icin hazirlanirlar. Bu kati yakitli roketlerin kalkistaki agirligi, yaklasik 580 tondur ve 11.800.000 newton'luk bir itme meydana getirmektedir. Uzunlugu 45.5 metre, silindirik govdenin capi ise 3.7 metredir. Uzay gemisinin ana motorlarina yakit veren buyuk dis tank ise yerden 200 km. Yukseklikte iken yakiti bittiginde aractan ayrilir. 20 katli bir apartman yuksekliginde (50 m.) Olan bu buyuk silindirik tankin capi 30 metredir.

Yapimi icin 30 ton aluminyum kullanilan bu tankin bir kez kullanilmasi, bir cok kisinin NASA'yi elestirmesine neden olmaktadir. Cunku mekikten ayrilan tank, daha sonra dunya atmosferine girerek yanmaktadir. NASA muhendisleri bu tanklardan nasil yararlanacaklarini dusunmektedirler. Hazirlanan bu projeye gore, 1990'dan sonra kurulmasi beklenen uzay istasyonunun, bu tanklardan yirmisinin bir araya getirilerek yapilmasi onerilmaktedir. Martin Marietta Aeorospace sirketi'nin gelistirilmis programlar baskani olan Frank Williams'a gore gemi, tankini uzayda biraz daha sonra birakacak. O zaman tank, yer atmosferine dusmeyecek, gemiyi izleyerek istenen yorungeye oturtulmasi saglanacak. Deneylerin yapilacagi ve incinde rahatca yasanilabilecek saglamlikta olan bu silindirler uc uca eklendiginde, istenen uzay istasyonunun hem daha kisa zamanda, hem de daha ekonomik bir sekiled yapilabilecegi ileri surulur. Uzay gemisinin on govdesi ve murettebat bolumu, aluminyumdan yapilmis uc kattan olusmaktadir. En ust katta, yorunge aracinin dendisini, tum uzay gemisi sistemini ve tasinan yuku yoneten, deneteleyen kumanda sistemi yer almaktadir. Bu katta, uc astronot iskemlesi bulunmaktadir. Orta kat, ucus zamani tasima ve yasam bolumu olarak ayrilmistir. Ayrica bu bolum, geminin yuk tasiyan dargo bolumu ile baglantilidir. Alt katta ise cevre kontrol gerecleri yer almaktadir. Geminin orta bolumu, yuk tasiyan kargo bulumudur ve uzaya giderken ustten acilan iki kapak ile ortulmektedir. Uzayda bu kapaklar acilarak, uydulari yorungeye oturtmak, yuruyus yapmak gibi cesitli gorevler yerine getirilmektedir. Arka govde ve motor yuvalarini tasiyan son bolum, yorunge aracinin en karmasik parcasidir. Sadece 8 dakika sureyle ateslenen ve yorungeye erismezden once 6 milyon newton'luk firlatma kuvveti yaratan uc ana motor bu bolumdedir. Ana motorlar sustuktan sonra gemiyi yorungesine oturtan iki roketten olusan yorunge manevra sistemi de bu arka bolumdedir. Son olarak bu bolumde 38'I ana, 6'si duyarli olmak uzere toplam 44 kucuk roketten olusmus, tepki - denetim sistemi vulunmaktadir. Bu sistem, aracin (yorunge icinde kalma kosulu ise) konumu ve uc ekseni boyunca donme hareketleri saglamaktadir. Yukarida kisaca ozelliklerini tanitmaya calistig iniz uzay gemisi ilk uzay ucusunu, 3 yillik gecikmeden sonra, 1981 yilinda yapti. Ucusa hazirlanan 4 uzay gemisinden ilk yapilani, Colombia adini tasiyordu. Ucus komutani ve pilot, ilk gemi seyrinin personeliydiler. 12 nisan 1981 Colombia Florida'daki firlatma ussunden havalandi. Dunya ceversinde 36 tur atan gemi kalkistan 54.5 saat sonra, 14 nisan gunu yeryuzune dondu. Ucus basarili gecmisti ama; gemiyi yuksek sicaktan koruyan fayanslari onemli derecede hasara ugramisti. Hasar nedeni olan sicaklik, ozellikle arac dunya'ya

donerken, atmosferdeki surtunmeden daynaklaniyordu. Ikinci ucus, 14 kasin 1981 gunu gerceklestirildi. Bes gun olarak dusunulen ucus programi yarida desildi ve gemi iki gun sonra yeryuzu'ne dondu. Bu ucusunda hava kirliligi, deniz arastirmalari gibi bir takim bilimsel arastirmalar yapildi. Ayrica, kanadalilarin yaptigi herhangi bir yone dogru 15.6 metre uzanabilen, gemi disindaki bir nesneyi tutmak icin veya icindeki bir aleti tutup uzaya birakabilmek icin kullanabilecek, kiminin vinc, kiminin robot, bazilarinin da mekanik kol dedigi birimi denediler. Bu ucusta gemi, birinciye gore daha az hasara ugramisti. Ucuncu ucus, 22 mart 1982 gunu basladi ve ilk kez sekiz gun surdu. Gemi, planlanan seyrini bir gun gecikmeyle 30 mart'ta tamamladi. Bu seyirde, komutan ve pilot, normal calismalarin yani sira, bir cok seyle de ugrastilar. Bunlar uzay tutmasi, radyo arizalari, tikanmis tuvalet, lumbuzlardaki kiragi, arizali radar ekrani ve uykusuzluktu. Fakat herseye karsin, cok basarili bir seyirdi. Astronotlar, geminin sadece bir yuzunu daima gunes'e cevirerek birkac saat isittilar, dogal olarak diger taraf da dondu. Boylece geminin isisal ozellikleri saptanmis oldu. Mekanik kola yerlestirilen bir cihazla, uzay gemisi cevresindeki parcaciklar ve elektrik alanlari olculdu. Mekanik kolun hareketini surekli denetim altinda tutmak icin kol uzerine yerlestirilen televizyon kamerasi arizalanica, personel ayni isi yapabilmek icin bildigimiz avci durbunu kullanmak zorunda kaldilar. Ilk ucus gununun sonunda, yeryuzu'nden havalanirken lumbuz koruyucusunu kiran beyaz maddenin, geminin bas kismindan kopan isi koruyucu oldugunu kesfettiler. Personel ilk gun hicbir sey yiyemedi. Ayrica pilot, agirliksiz ortama alisamadigindan uyuyamadi; dolayisiyla da ikinci gun cok yorgun dusmustu. Bu durumu pilot su sozlerle dile getiriyordu: "kendimi, sanki her on dakikada bir maraton kosuyormus gibi hissettim." Bu seyirde ayrica ari, pervane, ve sineklerden olusan hayvanlarin, agirliksiz ortamda davranislari incelendi. Arilar ucmaktan yorulduklarinda, amacsiz bir sekilde olduklari yere donuyorlardi. Gemi dunya'ya dondugunde tum arilar olmustu. Pervaneler cilgin bir sekilde kanat cirptilar; sinekler hep yuruduler. Pilot ucmak icin calisan bir sinegi asla gormedigini soyluyordu. Inisin yapilacagi Edwards hava kuvvetleri ussu'ndeki kuru gol yatagi mevsimin de etkisiyle inis gunu iyice islanmisti. Bu nedenle, inis oraya degil de, New Mexico'daki limana yapildi. Fakat inisin yapilacagi gun kuvvetli bir firtina patlamis ve inisin yapilacagi alan, seyirdeki gemiden dahi rahatca gorulebilinen beyaz bir toz bulutu altinda kalmisti. Bu nedenle ucus bir gun geciktirildi. Dorduncu ucus, 27 haziran - 4 temmuz 1982 arasi gerceklestirildi. Bu seyir digerlerinden iki yonden farkliydi. Birincisi, askeri amacli yuk tasiyordu. Hava kuvvetleri yukun ne oldugunu aciklamadi.

Fakat bu gizli yukun, kirmiziotesi arama ve tarama yapan bir alet oldugu biliniyordu. Ikinci farkli yon, ogrencilerin hazirladigi 90 kg. Agiriligindaki deney paketinin tasinmasiydi. Bu seyirde yapilan bir baska deney de bazi biyolojik materyalin birbirlerinden ayrilmasiydi. Deneyi yapan alet, bu materyal karisimi bir elektrik alana koyuyor ve onlari dogal elektrik yuklerine gore secebiliyordu. Dunya ustunde bu islemi, yercekimi etkilemekte elektrik yuku, kicaklik ve calkantiya neden olmakta, dolayisiyla da materyal tekrar birbirine karismaktadir. Uzayda bu materyalleri birbirinden ayirmanin, 800 kez daha etkin oldugu ortaya cikarildi. Bu son deneme ucusuydu. Bundan sonraki ucuslar, normal ticari amacli olacakti. Dorduncu ucusta basariya ulasamayan en onemli nokta, kati yakitli roketlerin parasut mekanizmasinin arizalanmasi ve her biri 7 milyar tl'na mal olan bu roketlerin deniz dibini boylamasiydi. Besinci ucusun personel sayisi, ilk kez ikiden fazla oluyordu. Ucus komutani ve pilottan baska, William ve Joseph adli iki astronot da ucus uzmani olarak gemide yer aldilar. Geminin ilk ticari yuku olan iletisim uydulari 11 kasim 1982 gunu baslayan bu seferde basariyla yorungeye oturtuldu. Eger bu uydular yerden yorungete yerlestirilseydi, uydu sahipleri daha fazla para odemek zorunda kalacaklardi. Bu seyirde personeli uzay tuttu. Bu yuzden uzayda yuruyus izlencesi bir gun ertelendi. Ertesi gun ise her biri yarim milyar tl'na mal olan uzay melbusati arizalandi. Tum ugraslara karsin arizalar giderilemedigi icin yuruyusten vazgecildi. Fakat bu cok onemli bir deneydi; cunki gelecekte uzay limani gibi buyuk yapilar insa edilirken, bu techizat ile arac disi calismalar yapilacak.

APPENDIX B
THE LISP PROGRAM

```
(denfun huffman (P)
(sortcar  (asign (arrange (mapcar 'list P))) 'greaterp))
(denfun arrange (Q)
(cond ((null  (cdr Q))  Q)
(t (arrage (insert (list (add (caar Q)  (caadr Q))
(car Q)   (cadr Q))
(denfun insert (x Q)
(cond ((null Q) (cons x Q)
((lessp (plus (car x) epsilon) (caar Q)) (putin N x Q))
(t (cons (car Q)  (insert x (cdr Q)) )) ))

(denfun putin (n x L)
(cond ((zerop n ) (cons x L))
((null L ) (list x))
(t  (cons (car L) (putin (subi n) x (cdr L))))))
(defun assign (Q0 (split nil (car Q)) 0

(defun split (c l)
(cond ((null (cdr L)) (list ( list (car L) c)) )
(t (append (split (cons 1 c) (cadr L))
(split (cons 0 c) (cadr  L)) )) ))

(defun sortcode (L)
(cond ((null L) nil)
(t (inscode (caar L) (cadar L) (sortcode (cdr L)) )) ))
(defun inscode (p c L)
(cond ((null L) (list (list p c)) )
((greaterp (length c) (length (cadar L)))
(cons (list p (cadar L)) (inscode (caar L) c (cdr L)) ))

(defun totlength (L)
(cond ((null L) 0)
```

```
(t (add (times (caar L) (length (cadar L)) )
(totlength (cdr L)) )) ))

(defun avglength (L)
(qoutient (times 1.0 (totlength L))
(apply 'add (mapcar 'car L)) ))

(defun varlength (L)
(quotient (times 1.0 (varlength2 L (avglength L)))
(apply 'add (mapcar 'car L))))

(defun varlength2 (L mu)
(cond ((null L) 0)
(t (add (times (caar L)
(expt (differnce  (length(cadar L)) mu) 2))
(varlength2 (cdr L) mu)))))

(defun Zipf (n)
(cond ((zerop n) nil)
(t (cons (quotient 1.0 n) (Zipf (- n 1)) )) ))


(defun tryN (n e)
(set 'N n)
(set 'epsilon e)
(set 'code (sortcode (huffman Turkish)) )
(print (list 'N '=  n 'epsilon '= e))
(pp code)
(print (list 'mean '=  (avglength code))) (terpr)
(print (list 'variance '=  (varlength code))) (terpr))
(set 'Prob '(20 25 33 50))


(set 'Turkish
'(0 6 6 17 28 34 39 45 45 56
61 67 67 73 73 84 84 89 112 134
162 196 358 581 687 872 989 1017 1224 1637
```

```
1883 2185 2660 2682 2945 3213 3509 3861 3984 5130
5163 6085 6611 7952 9427 10528 13339))

(set 'N 0)
(set 'epsilon 0)
```

APPENDIX C
PASCAL COMPUTER PROGRAM WHICH DROPS THE SYMBOLS

```
    Program Somethihg (INPUT, OUTPUT)
    VAR


    ch : CHAR
    x
INTEGER;



    BEGIN
    x:= 0
    WHILE not EOF DO



    Begin
    READ (ch )
    IF ( ch = ' any symbol to be dropped')
    x =: x + 1


    ELSE
    WRITE (ch)
    End


    WRITELN
    WRITELN: (' The Number of the Dropped Symbols is ', x)


    END.
```

APPENDIX D
SHORT TURKISH ARTICLES

1. THE FIRST ARTICLE

Genel Bilgiler : 1. Yabancilar ve yurt disinda calisan Turkler girislerinde beyan etmek kosuluyla 3,000Amerikan dolari veya esitini asan dovileri beraberlerinde yurt disina cikarabilirler. 2. Yolcular encok 1,000 Amerikan dolari karsiligi Turk parasini yurt disina cikarabilirler. 3. Yolcular kendilerine ait 3,00 Amerikan dolarini asmayan ziynet esyalarini giriste beyan edilmek sarti ile yurt disina goturebilirler. 4. Yolcular degerine bakilmaksizin, gumruk mevzuatina uygun, sahsi, ailevi, mesleki ve trustik nitelikteki esyalari beraberlerinde goturebilirler. [Ref. 9]

2. THE SECOND ARTICLE

Kisisel Esya : 1. Yolcunun giyinip kusanmasina, kullanmasina, suslenmesine ait (ic camasirlair, gomlek, kravat, elbise, palto, manto, sapka, ayakkabi, toka dugme, kupe, bilezik, yuzuk, birer adet cep ve kol saati, medil, corap, pijama, perdesu, semsiye gibi esya ile yurt disinada iki yil veya daha fazla kalip Turkiye'ye kesin donen kisinin bir adet kurkten mamul giyim esyasi). 2. Yolcunun okumasina ve yazmasina ait esya (kitap, dergi, kursun kalem, kagit, defter, kristal, gumus veya kiymetli madenlerden olanalr haric yazi takimi gibi). 3. Bir adet portatif yazi makinasi. [Ref. 9]

3. THE THIRD ARTICLE

Tip bilimi, kadin ile erkek arasinda bir ustunluk sorunu degil, sadece bir "farklilik" oldugunu soyluyor. 18 yasindaki bir kadin (yada bir kiz), ayni yastaki bir erkekten ortalama 10 santimetre daha kisa ve gene ortalama 13 kilo daha hafif. Erkeklerde adele yapisi vucud agirliginin yuzde kirk kadarini olustururken, bu oran kadinlarda yuzde 23 dolayinda kaliyor. Bir baska acidan bakilirsa, kadin vucudunda erkege nazaran yuzde 10 kadar fazla yag dokusu var. Ve, hepsi bu. Bu farkliliklar, kadinin dayaniklilik gerektiren sporlarda erkeklere gore daha avantajli olmalarini sagliyor. Buna karsilik erkeklerde adele gucu isteyen sporlarda daha basarili.Maratonda erkek ile kadin arasindaki der ece farklarinin 2,000 yili civarinda ortadan kalkacagi saniliyor. Mans'i gecerken en iyi 10 derece yapmis olan yuzuculerden 8'i kadin. [Ref. 10]

4. THE FOURTH ARTICLE

Maraton parkuru tam 42,195 metre tutuyor. Kosunun asagi yukari 10. kilometresinde, vucud, sinirlerinden erdorfin denilen bir madde salgilayarak kan dolasimina katmaya basliyor. Endorfin, bunyenin olusturdugu dogal bir uysturucu. Gorevide, vucud biolojik bir olum kalim savasi verirken aci duygusunu onleyip mucadelenin surmesini saglamak. Tip biliminin elde ettigi bulgulara gore, maraton kosucularinin yuzde 3 kadari, bir sure sonra bu maddenin kesin tutkunu haline geliyorlar. Kosmadiklari taktirde, tipki eroin muptelalari gibi, uykulari kaciyor, saldirgan bir tavir aliyorlar, hatta bunalim, korku gibi surekli ruh bozukluklari gosterenlere bile rastlaniyor. Digerlerinde de, aci hissi ortadan kalktigindan dolayi, "kendini cok iyi hissetme" hali gozleniyor. [Ref. 11]

APPENDIX E
REWRITTEN ARTICLES AT STEP 6

1. THE FIRST ARTICLE

Genel bilgiler : 1 Yabancilar e yurt disinda calisan Turkler girislerinde beyan etmek kouluyla 3000 Amerikan dolari eya esitini asmayan doileri beraberlerinde yurt disina cikarabilirler 2 Yolcular encok 1000 Amerikan dolari karsiligi Turk arasini yurt disina cikarabilirler 3 Yolcular kendilerine ait 3000 Amerikan dolarini asmayan iynet esyalarini beraberlerinde yurda getirebilirler veya yurt disina cikarabilirler Kiymeti 3000 Amerikan dolarindan yukari olan iynet esyalari giriste beyan edilmek sarti ile yurt disina goturulebilir 4 Yolcular sasi, ailei, mesleki e turistik nitelikteki esyalari beraberlerinde goturebilirler

2. THE SECOND ARTCLE

Kisisel esya : 1 Yolcunun giyini kusanmasina, suslenmesine ait esya ic camasirlari, gomlek, kraat, elbise, alto, manto, saka, ayyakkabi, toka dugme, kue, bileik, yuuk, birer adet ce e kol saati, mendil, cora, iama, erdusu, semsiye gibi esya ile yurt disinda iki yil eya daa ala kali Turkiye'ye kesin donen kisinin bir adet kurkten mamul giyimesyasi 2 Yolcunun okumasina e yamasina ait esya kita dergi, kursun e murekkeli kalem, kagit, deter, kristal, gumus e kiymetli madenlerden olanalar aric yai takimi gibi 3 bir ade ortati yai makinasi

3. THE THIRD ARTICLE

Ti bilimi, kadin ile erkek arasinda bir ustunluk sorunu degil, sadece bir arklilik oldugunu soyluyor 18 yasindaki bir kadin yada ki, ayni yastaki bir erkekte ortalama 10 santimetre daa kisa e gene ortalama 13 kilo daa ai Erkeklerde adele yaisi ucud agirligini yude 40 kadarini olustururken, bu oran kadinlarda yude 23 dolayinda kaliyor Baska bir acidan bakilirsa, kadin ucudunda erkege nazaran yude 10 kadar ala yag dukusu ar e, esi bu Bu arkliliklar, kadinin dayaniklilik gerektiren sorlarda erkeklere gore daa aantali olmalarini sagliyor Buna karsilik erkeklerde adele gucu isteyen sorlarda daa basarili Maratonda erkek ile kadin arasindaki derece arklarinin 2000 yili ciarinda ortadan kalkacagi saniliyor Mansi gecen en iyi 10 derece yamis olan yuuculerden 8 i kadin

4. THE FOURTH ARTICLE

Maraton arkuru tam 42195 metre tutuyor Kosunun asgi yukari 10 kilometresinde, ucud, sinirlerinden endorin denilen bir madde salgilayarak kan dolasimina katmaya basliyor Endorin,bunyenin olusturdugu dogal bir uyusturucu Goreide, ucud biolojik bir olum kalim saasi erirken aci duygusunu onleyi mucadelenin surmesini saglamak Ti bilimin elde ettigi bulgulara gore, maraton kosucularinin yude 3 kadari, bir sure sonra bu maddenin kesin tutkunu aline geliyorlar Kosmadiklari taktirde, tiki eoin mutelelari gibi, uykulari kaciyor, saldirgan bir tair aliyorlar atta bunalim, korku gibi surekli ru boukluklari gosterenlere bile rastlaniyor Digerlerinde de aci issi ortadan kalktigindan dolayi, kendini cok iyi issetme ali goleniyor

The Number of the Dropped Symbols is : 140

APPENDIX F
SIMULATION PROGRAM WITHOUT DROPPING PROCESS

```
Program Buffersize (input, output)
Var
CH:  Char
X:  integer
MAXBUF, BUF1, INPUTR, OUTR, LE, BUF:  real
BEGIN
LE:= 0.0
BUF := 0.0
MAXBUF:= 0.0
BUF1 :  = 1.0;
x:=0
INPUTR := 6.0
OUTR := 4.01359
WHILE not EOF DO
BEGIN
READ (ch)
x:= x+1
IF (ch='I') THEN
LE :=
ELSE IF (ch='A') THEN
LE :=
ELSE IF (ch='E') THEN
LE :=
ELSE IF (ch='N') THEN
LE :=
ELSE IF (ch='R') THEN
LE :=
ELSE IF (ch='U') THEN
LE :=
ELSE IF (ch='L') THEN
```

```
LE :=
ELSE IF (ch='S') THEN
LE :=
ELSE IF (ch='K') THEN
LE :=
ELSE IF (ch='D') THEN
LE :=
ELSE IF (ch='T') THEN
LE :=

ELSE IF (ch='M') THEN
LE :=
ELSE IF (ch='Y') THEN
LE :=
ELSE IF (ch='O') THEN
LE :=
ELSE IF (ch='G') THEN
LE :=
ELSE IF (ch='B') THEN
LE :=
ELSE IF (ch='C') THEN
LE :=
ELSE IF (ch=',') THEN
LE :=
ELSE IF (ch='.') THEN
LE :=
ELSE IF (ch='Z') THEN
LE :=
ELSE IF (ch='V') THEN
LE :=
ELSE IF (ch='P') THEN
LE :=
ELSE IF (ch='H') THEN
LE :=
ELSE IF (ch='F') THEN
```

```
LE :=
ELSE IF (ch='0') THEN
LE :=
ELSE IF (ch=''') THEN
LE :=
ELSE IF (ch='1') THEN
LE :=
ELSE IF (ch='"') THEN
LE :=
ELSE IF (ch='2') THEN
LE :=
ELSE IF (ch=')') THEN
LE :=
ELSE IF (ch='5') THEN
LE :=
ELSE IF (ch='3') THEN
LE :=
ELSE IF (ch='8') THEN
LE :=
ELSE IF (ch='(') THEN
LE :=
ELSE IF (ch='4') THEN
LE :=
ELSE IF (ch=' ') THEN
LE :=
ELSE IF (ch='9') THEN
LE :=
ELSE IF (ch='J') THEN
LE :=
ELSE IF (ch='6') THEN
LE :=
ELSE IF (ch='W') THEN
LE :=
ELSE IF (ch=':') THEN
```

```
LE :=
ELSE IF (ch='7') THEN
LE :=
ELSE IF (ch='-') THEN
LE :=
ELSE IF (ch='?') THEN
LE :=
ELSE IF (ch='X') THEN
LE :=
ELSE IF (ch='Q') THEN
LE :=
BUF := BUF + LE
BUF  := BUF - OUTR
IF ( BUF < 0.0 ) THEN
BUF := 0.0
BUF1 := BUF
IF (MAXBUF < BUF1 ) THEN
MAXBUF := BUF1
ELSE
MAXBUF := MAXBUF
END
END
Z := X - Y
WRITELN ( ' BUFFER     ', BUF)
WRITELN (' REQUIRED BUFFER SIZE FOR', X, ' CHARACTERS IS',
MAXBUF)
WRITELN (' OUTPUT RATE IS', OUTR)
WRITELN (' INPUT RATE IS ', INPUTR)
END.
SENTRY
(  MESSAGE  )
```

APPENDIX G
SIMULATION PROGRAM WITH DROPPING PROCESS

```
Program Buffersize (input, output)
Var
CH:  Char
Z, Y, X:  integer
MAXBUF, BUF1, INPUTR, OUTR, LE, BUF:  real
BEGIN
LE:= 0.0
BUF := 0.0
MAXBUF:= 0.0
BUF1 :  = 1.0;
x:=0
z:= 0
y := 0
INPUTR := 6.0
OUTR := 4.01359
WHILE not EOF DO
BEGIN
READ (ch)
IF (ch='Q') or (ch='X') or (ch='?') or (ch='-') or (ch=':') or
(ch='J') or (ch=' ') or (ch='(') or (ch=')') or (ch='"') or
(ch=''') or (ch='F') or (ch='H') or (ch='P') or (ch='V') or
(ch='Z') or (ch='.') or (ch=',') or (ch='W') THEN
x:= x+1
ELSE
BEGIN
IF (ch='I') THEN
LE :=
ELSE IF (ch='A') THEN
LE :=
ELSE IF (ch='I') THEN
```

```
LE :=
ELSE IF (ch='E') THEN
LE :=
ELSE IF (ch='N') THEN
LE :=
ELSE IF (ch='R') THEN
LE :=
ELSE IF (ch='U') THEN
LE :=
ELSE IF (ch='L') THEN
LE :=
ELSE IF (ch='S') THEN
LE :=
ELSE IF (ch='K') THEN
LE :=
ELSE IF (ch='D') THEN
LE :=
ELSE IF (ch='T') THEN
LE :=

ELSE IF (ch='M') THEN
LE :=
ELSE IF (ch='Y') THEN
LE :=
ELSE IF (ch='0') THEN
LE :=
ELSE IF (ch='G') THEN
LE :=
ELSE IF (ch='B') THEN
LE :=
ELSE IF (ch='0') THEN
LE :=
ELSE IF (ch='1') THEN
LE :=
ELSE IF (ch='2') THEN
```

```
LE :=
ELSE IF (ch='3') THEN
LE :=
ELSE IF (ch='5') THEN
LE :=
ELSE IF (ch='8') THEN
LE :=
ELSE IF (ch='4') THEN
LE :=
ELSE IF (ch='9') THEN
LE :=
ELSE IF (ch='6') THEN
LE :=
ELSE IF (ch='7') THEN
LE :=
BUF := BUF + LE
BUF  := BUF - OUTR
IF ( BUF < 0.0 ) THEN
BUF := 0.0
BUF1 := BUF
IF (MAXBUF < BUF1 ) THEN
MAXBUF := BUF1
ELSE
MAXBUF := MAXBUF
END
END
Z := X - Y
WRITELN ( ' BUFFER     ', BUF)
WRITELN (' REQUIRED BUFFER SIZE FOR', Z, ' CHARACTERS IS',
MAXBUF)
WRITELN (' TOTAL NUMBER OF CHARACTERS IS', X )
WRITELN (' NUMBER OF DROPPED SYMBOLS IS', Y)
WRITELN (' OUTPUT RATE IS', OUTR)
WRITELN (' INPUT RATE IS ', INPUTR)
```

END.
SENTRY
(MESSAGE)

APPENDIX H
REWRITTEN ARTICLES AFTER DROPPING I AND N

1. THE FIRST ARTICLE

geel blgler : 1. Yabaclar ve yurt dsda calsa Turkler grslerde beya etmek kouluyla 3000 Amerka dolar vaya est asmayan dovler beraberlerde yurt dsa ckarablrler. 2. Yolcular ecok 1000 Amerka dolar karslg Turk paras yurt dsa ckarablrler. 3. Yolcular kedlere at 3000 Amerka dolarda yukar ola zyet esyalar grste beya edlmek sart le yurt dsa goturuleblr. 4. Yolcular, sahs, alev, meslek ve turustk telktek esyalar beraberlerde beraberlerde gotureblrler.

2. THE SECOND ARTICLE

Kssel Esya : 1. Yolcuu gyp kusamasa, suslemese at esya (c camasrlar gomlek, kravat, elbse, palto, mato, sapka, ayyakkab, toka, dugme, kupe, blezk, yuzuk, brer adet cep ve kol saat, medl, corap, pjama, perdesu, semsye gb esya le yurt dsda k yl veya daha fazla kalp Turkye'ye kes doe ks br adet kurkte mamul gym esyas) 2. Yolcuu okumsasa ve yazmasa at esya (ktap, derg, kursu ve murekkepl kalem, kagt, defter, krstal, gumus ve kymetl maddelerde olalar harc yaz takm gb) 3. Br adet portatf yaz makas.

3. THE THIRD ARTICLE

Tp blm, kad le erkek arasda br ustuluk soruu degl, sadece br "farkllk" olduguu soyluyor. 18 yasdak br kad (yada kz), ay yastak br erkekte ortalama 10 satimetere daha ksa ve gene ortalama 13 klo daha haff Erkeklerde adele yaps vucud agrlg yuzde krk kadar olustururke, bu ora kadlarda

yuzde 23 dolayda kalyor. Baska br acdan baklrsa, kad vucududa erkege nazara yuzde 10 fazla yag dokusu var. Ve, heps bu. Bu farkllklar, kad dayakllk gerektre sporlarda erkeklere gore daha avatajl olmalar saglyor. Bua karslk erkeklerde adle gucu steye sporlarda daha basarl. Maratoda erke le kad arasdak derece farklar 2000 yl cvarda ortada kalkacag salyor. Mas gece e y 10 derece yapms ola yuzuculerde 8 i kad.

4. THE THIRD ARTICLE

Marato parkuru tam 42,195 metre tutuyor. Kosuu asg yukar 10. klometresde, vucud slerde edorf dele br madde salglayarak ka dolasma katmaya baslyor. Edof, buye

olusturdugu dogalbr uysturucu. Gorede, vucud bolojk br olum kalm savas verrke ac duygusuu oleyp mucadele surmes saglamak. Tp blm elde ettg bulgulara gore, marato kosucular yuzde 3 kadar, br sure sora bu madde kes tutkuu hale gelyorlar. Kosmadklar taktrd tpk ero muptelalar gb, uykular kacyor, saldrga br tavr alyorlar. Hatta bualm, korku gb surekl ruh bozukluklar gosterelere ble rastlayor. Digerlerde de, ac hss ortada kalktgda dolay "kend cok y hssetme" hal gozleyor.

The Number of the Dropped Symbols is : 448

APPENDIX I

REWRITTEN ARTICLES AFTER DROPPING THE SYMBOLS COMBINATION

1. THE FIRST ARTICLE

geel blgler 1 Yabaclar ve yurt dsda calsa Turkler grslerde beya etmek kouluyla 3000 Amerka dolar vaya est asmayan dovler beraberlerde yurt dsa ckarablrler 2 Yolcular ecok 1000 Amerka dolar karslg Turk paras yurt dsa ckarablrler 3 Yolcular kedlere at 3000 Amerka dolarda yukar ola zyet esyalar grste beya edlmek sart le yurt dsa goturuleblr 4 Yolcular sahs alev meslek ve turustk telktek esyalar beraberlerde beraberlerde gotureblrler

2. THE SECOND ARTICLE

Kssel Esya 1 Yolcuu gyp kusamasa suslemese at esya c camasrlar gomlek kravat elbse palto mato sapka ayyakkab toka dugme kupe blezk yuzuk brer adet cep ve kol saat medl corap pjama perdesu semsye gb esya le yurt dsda k yl veya daha fazla kalp Turkyeye kes doe ks br adet kurkte mamul gym esyas 2 Yolcuu okumsasa ve yazmasa at esya ktap derg kursu ve murekkepl kalem kagt defter krstal gumus ve kymetl maddelerde olalar harc yaz takm gb 3 Br adet portatf yaz makas

3. THE THIRD ARTICLE

Tp blm kad le erkek arasda br ustuluk soruu degl sadece br farkllk olduguu soyluyor 18 yasdak br kad yada kz ay yastak br erkekte ortalama 10 satimetere daha ksa ve gene ortalama 13 klo daha haff Erkeklerde adele yaps vucud agrlg yuzde krk kadar olustururke bu ora kad a yuzde 23 dolayda kalyor Baska br acdan baklrsa kad vucududa erkege nazara yuzde 10 fazla yag dokusu var Ve heps bu Bu farkllklar kad dayakllk gerektre sporlarda erkeklere gore daha avatajl olmalar saglyor Bua karslk erkeklerde adle gucu steye sporlarda daha basarl Maratoda erke le kad arasdak derece farklar 2000 yl cvarda ortada kalkacag salyor Mas gece e y 10 derece yapms ola yuzuculerde 8 i kad

4. THE THIRD ARTICLE

Marato parkuru tam 42195 metre tutuyor Kosuu asg yukar 10 klometresde vucud slerde edorf dele br madde salglayarak ka dolasma katmaya baslyor Edof buye olusturdugu dogal br uysturucu Gorede vucud bolojk br olum kalm savas verrke ac

duygusuu oleyp mucadele surmes saglamak Tp blm elde ettg bulgulara gore marato kosucular yuzde 3 kadar br sure sora bu madde kes tutkuu hale gelyorlar Kosmadklar taktrd

tpk ero muptelalar gb uykular kacyor saldrga br tavr alyorlar Hatta bualm korku gb surekl ruh bozukluklar gosterelere ble rastlayor Digerlerde de ac hss ortada kalktgda dolay kend cok y hssetme hal gozleyor

The number of the dropped symbols is 544

LIST OF REFERENCES

1. Kilic, S., *Modification Of Huffman Coding*, Master's Thesis, Naval Postgraduate School, Monterey, CA, March 1985.

2. Akinsel, S., *Reduction In Bandwidth By Using Variable Length Codes*, Master's Thesis, Naval Postgraduate School, Montery, CA, December 1985.

3. Feher,K., *Digital Communications Microwave Applications*, Prentice-Hall,1981.

4. Alisoukos, V. F. and Tumasi, W., *Digital and Data Communications*, Prentice-Hall, 1985.

5. Bellamy, J., *Digital Telephony*, John Willey & Sons, 1982.

6. Hamming, R. W., *Coding And Information Theory*, Prentice-Hall, Inc., 1980.

7. Huffman, D., "A Method For The Construction Of The Minimum Redundacy Codes," *Proceeding Of The Institude Of The Radio Engineers*, pp. 1098-1101, Vol. 40., September 1952.

8. Turban, E. & Meredit, J.R. *Fundamentals of Management Science* Bussiness Puplications, INC., 1985.

9. Garanti Bankasi *Buyuk Gumruk Rehberi*, Pg. 5-6.

10. Nokta, Weekly Turkish Magazine, *Kadin Ve Spor/ Turkiye'de Durum*, Pg. 76, Vol. 44, November, 1986.

11. Nokta, Weekly Turkish Magazine *Maraton/Tutkunun Sirri* Pg. 76, Vol. 46, November 1986.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Deniz Kuvvetleri Komutanligi Bakanliklar, Ankara / Turkey	5
4.	Deniz Harp Okulu Komutanligi Kutuphanesi Tuzla, Istanbul / Turkey	1
5.	Department Chairman, Code 54Gk Administration Science Naval Postgraduate School Monterey, California 93943-5000	1
6.	Professor Richard W. Hamming, Code 52Hg Computer Science Department Naval Postgraduate School Monterey, California 93943-5000	3
7.	Professor Daniel R. Dolk, Code 54Dk Administartion Science Department Naval Postgraduate School Monterey, California 93943-5000	3
8.	Ltjg, Ahmet Corapcioglu 1'nci Orta Sokak, Osmanli Apt., No:47/16 Goztepe, Istanbul / Turkey	4

Thesis
C754545 Corapcioglu
c.1
Reduction in bandwidth and buffer size by using modified Huffman coding after dropping the less frequent source symbols.